NAVAL POSTGRADUATE SCHOOL
MONTEREY CALIFORNIA 93940

NPS55-77-7

NAVAL POSTGRADUATE SCHOOL

Monterey, California

ASYMPTOTIC EFFICIENCY AND

SOME QUASI-METHOD OF MOMENTS ESTIMATORS

Robert R. Read

March 1977

Approved for public release; distribution unlimited

FEDDOCS
D 208.14/2:NPS-55-77-7

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPS-55-77-7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) ASYMPTOTIC EFFICIENCY AND SOME QUASI-METHOD OF MOMENTS ESTIMATORS		5. TYPE OF REPORT & PERIOD COVERED Techical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(*s*) Robert R. Read		8. CONTRACT OR GRANT NUMBER(*s*)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE 15 March 1977
		13. NUMBER OF PAGES 69
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

Estimation
Method of Moments
Maximum Likelihood
Efficiency

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

The report contains the asymptotic efficiencies of some candidate estimators which provide alternatives to maximum likelihood in some common probabilistic settings. The alternative estimators can be found with measurably less effort than solving the likelihood equations. They include the method of moments and similarly constructed estimators that involve the harmonic mean. The most successful example found deals with the negative binomial distribution. Here, the harmonic mean estimator has high efficiency in regions where the method of moments estimator has rather low efficiency.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

ASYMPTOTIC EFFICIENCY AND

SOME QUASI-METHOD OF MOMENTS ESTIMATORS

R. R. Read
Naval Postgraduate School
Monterey, Ca. 93940

ABSTRACT

The report contains the asymptotic efficiencies of some candidate estimators which provide alternatives to maximum likelihood in some common probabilistic settings. The alternative estimators can be found with measurably less effort than solving the likelihood equations. They include the method of moments and similarly constructed estimators that involve the harmonic mean. The most successful example found deals with the negative binomial distribution. Here, the harmonic mean estimator has high efficiency in regions where the method of moments estimator has rather low efficiency.

TABLE OF CONTENTS

I. Introduction

The need for readily computed parameter estimates is great. Maximum likelihood estimators are known to be asymptotically efficient, but often they are very hard to find. The most popular alternative is the method of moments which usually yields readily computed estimates, but sometimes these estimates are not very efficient. This report looks at the efficiency of the method of moments and of some similarly constructed 'quasi-method of moment' estimators.

The basic idea is to select a system of estimating equations which equates various statistics to their expected values. The method of moments does this for sample moments of order one, two, etc. We propose to consider also moments of order zero, minus one, and perhaps other functions. The examination of the consequent efficiencies may aid in the building of intuition so that a wiser selection of estimating functions can be made in new situations when they appear.

Moments of order zero and minus one require positive data. The former is the geometric mean and the latter is the harmonic mean. They form part of a general family, $f(r) = [\frac{1}{n} \sum_1^n x_i^r]^{1/r}$, of nondecreasing means (where $x_1, \ldots, x_n$ forms the data). When dealing with the harmonic mean, we will be setting $\frac{1}{n}\sum \frac{1}{x_i}$ $(= [f(-1)]^{-1})$ equal to $E(1/X)$ because, for the parent populations treated here, the latter value is easy to obtain. Similarly when dealing with the geometric mean, we will be setting $\frac{1}{n} \sum_1^n \ln x_i$ $(= \ln f(0))$ equal to $E(\ln X)$. In this form, the geometric mean appears in many maximum likelihood systems. This suggests that alternative quantities, that are close to means with $r = 0$, might be profitably exploited. The results give some support to this idea.

The structure of the efficiency computations utilizes the theoretical work presented in a companion report [12]. The pertinent material is

summarized in Section II of the present report. Section III contains applications of the idea to two, one-parameter parent populations, Poisson and symmetric beta. Some speculations for interpreting these results are made. Section IV is devoted to two parameter settings; gamma, negative binomial, and beta.

Much computational work is necessary to support this development and the details are relegated to the appendices. Appendix A contains general relationships among the required population moments. Specialization to Poisson, gamma, negative binomial, and beta is contained in Appendices B, C, D, E (resp.). Computations are performed by APL programs and these are included too, as Appendix F.

II. Methodology

It is assumed that the estimating equations have the form

$$g(x,\theta) = 0 \qquad (2.1)$$

where $x = (x_1,\dots,x_n)$ is the data vector of a random sample from the specifie parent population, $\theta' = (\theta_1,\dots,\theta_p)$ is the parameter point which belongs to an open subset of p-dimensional space, and $g' = (g_1,\dots,g_p)$. Primes denote transpose. In order for the system (2.1) to have a unique solution $\tilde{\theta}$, it is necessary (by the Implicit Function Theorem [16]) that the Jacobian $J = \left|\frac{\partial g_1,\dots,g_p}{\partial\theta_1,\dots,\theta_p}\right| \neq 0$.

The following structural assumptions are made:

(i) Each $g_j(x,\theta)$ for $j = 1,\dots,p$ is a symmetric function of x, i.e. is invariant under permutations of $x_1,\dots,x_n$.

(ii) $E\{g_j(X,\theta)\} = 0$ for $j = 1,\ldots,p$.

(iii) $\mathrm{Var}\{g_j(X,\theta)\} = \frac{1}{n} C_j(\theta) + o(\frac{1}{n})$ for $j = 1,\ldots,p$.

(iv) The $g_j(x,\theta)$ have bounded continuous partial derivatives with respect to $\theta_1,\ldots,\theta_p$ for $j = 1,\ldots,p$.

(v) $\tilde{\theta}$, which is the solution of (2.1) is consistent and asymptotically normal.

Assumption (i) is modest and expected of any reasonable estimating scheme. The meeting of assumptions (ii) and (iii) is a matter of scaling and arrangement. Assumption (iv) is needed so that the asymptotic covariance matrix is well behaved. Assumption (v) is always desirable and convenient since it implies that the estimators are asymptotically unbiased and the ellipsoid of concentration [4] is characterized in terms of the inverse of the asymptotic covariance matrix. Efficiency computations are based on its determinant.

Equipment for verifying (v) is contained in [9]. There, the functions g_j are averages of the form $\frac{1}{n}\sum_{i=1}^{n} g_j(x_i,\theta)$ for $j = 1,\ldots,p$ and this structure is consonant with the present set of assumptions. All of the cases treated in this report have this structure.

Much license is taken in what follows. The purist is referred to [9]. Let $A(x,\theta)$ be the p by p matrix of partial derivatives $\{\partial g_j/\partial\theta_k\}$. Assume that the elements behave as in (2.2)

$$A_{jk}(X,\tilde{\theta}) \to E\{\partial g_j/\partial\theta_k\} \qquad (2.2)$$

as $n \to \infty$. The resulting limiting matrix is denoted by A or $A(\theta)$. The assumptions allow the first order expansion

$$g(x,\theta) = g(x,\tilde{\theta}) + A(x,\tilde{\theta} + \rho(\theta-\tilde{\theta}))(\theta-\tilde{\theta}) \tag{2.3}$$

where ρ is a diagonal matrix of random numbers belonging to the interval [0,1]. Since the system is soluble, $g(x,\tilde{\theta}) = 0$ and we can write

$$(\tilde{\theta}-\theta) = -A^{-1}(x, \tilde{\theta} + \rho(\theta-\tilde{\theta})\ g(x,\theta)) \tag{2.4}$$

since the continuity of A implies that of A^{-1} and of g. Letting the asymptotic covariance matrices, as $n \to \infty$, be defined by

$$M = \text{limit } nE(\tilde{\theta}-\theta)(\tilde{\theta}-\theta)' \tag{2.5}$$

$$C = \text{limit } nE\{g(X,\theta)\ g'(X,\theta)\} \tag{2.6}$$

it follows from (2.2) and (2.4) that

$$M = A^{-1}\ C(A')^{-1} \tag{2.7}$$

The method of maximum likelihood fits into this setting. The parent population has density $f(x,\theta)$, and (2.1) takes the form

$$\frac{1}{n} S_r = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln f(x_i,\theta)}{\partial \theta_r} = 0 \qquad \text{for} \quad r = 1,\ldots,p \tag{2.8}$$

Assumption (ii) requires the regularity conditions [10] as does the relationship

$$E(S_r S_k) = -nE \left\{ \frac{\partial^2 \ln f(x,\theta)}{\partial\theta_r\ \partial\theta_k} \right\} = n\Lambda_{rk} \tag{2.9}$$

where Λ is the information matrix. Using (2.2) and (2.6) it is seen that both A and C are equal to Λ and hence

$$M = \Lambda^{-1} \tag{2.10}$$

follows from (2.7).

Now suppose that only the first q likelihood equations (2.8) are used in the system $g = 0$. Let us denote this subset by $\mu = 0$, and the remaining $p-q$ equations by $h = 0$. Thus in partitioned form (2.1) becomes

$$g = \left\{ \begin{matrix} \mu \\ h \end{matrix} \right\} = 0 \tag{2.11}$$

Proceeding formally, (2.6) becomes

$$C = \text{limit } n \begin{Bmatrix} E(\mu\mu') & E(\mu h') \\ E(h'\mu) & E(hh') \end{Bmatrix} = \begin{Bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Bmatrix} \tag{2.12}$$

The information matrix can be partitioned likewise

$$\Lambda = \begin{Bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{Bmatrix} \tag{2.13}$$

Further, define a $p-q$ by q matrix

$$g_{21} = \{E(\partial h_j/\partial\theta_k)\}, \qquad j = 1,\ldots,p-q;\ k = 1,\ldots,q \tag{2.14}$$

and a $p-q$ order square matrix

$$g_{22} = \{E(\partial h_j/\partial\theta_k)\}, \qquad j = 1,\ldots,p-q;\ k = q+1,\ldots,p \tag{2.15}$$

It is shown in [12, Sec. IV], that

$$C = \begin{Bmatrix} \Lambda_{11} & -g'_{21} \\ -g_{21} & C_{22} \end{Bmatrix} \tag{2.16}$$

where C_{22} is as (2.12),

$$A = \begin{Bmatrix} -\Lambda_{11} & -\Lambda_{12} \\ g_{21} & g_{22} \end{Bmatrix} \tag{2.17}$$

and

$$M^{-1} = \begin{Bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & H \end{Bmatrix} \qquad (2.18)$$

where

$$H = \Lambda_{21}G_{11}\Lambda_{12} - \Lambda_{21}G_{12}g_{22} - (\Lambda_{21}G_{12}g_{22})' + g_{22}G_{22}g_{22}$$

$$G_{11} = \{\Lambda_{11} - g_{12}C_{22}^{-1}g_{21}\}^{-1}$$

$$G_{12} = \Lambda_{11}^{-1}g_{21}'G_{22} \qquad (2.19)$$

$$G_{22} = \{C_{22} - g_{21}\Lambda_{11}^{-1}g_{21}'\}^{-1}$$

Because of (2.7), the determinant of (2.18) is

$$|M^{-1}| = |A|^2/|C| \qquad (2.20)$$

and it is useful to draw attention to its computation. Using partitioned form, see [6, p. 165], its ingredients have determinants

$$|C| = |\Lambda_{11}||C_{22} - g_{21}\Lambda_{11}^{-1}g_{12}|$$

$$|A| = |\Lambda_{11}||g_{22} - g_{21}\Lambda_{11}^{-1}\Lambda_{12}| \qquad (2.21)$$

and this is useful in computing the asymptotic efficiency, [12]

$$\mathrm{Eff}(\tilde{\theta}) = |M^{-1}|/|\Lambda| \qquad (2.22)$$

In the special case of $p = 2$ and $q = 1$, (2.19) reduces to

$$H = \frac{1}{|C|} \{\Lambda_{12}^2 C_{22} - 2\Lambda_{12} g_{21} g_{22} + \Lambda_{11} g_{22}^2\} \tag{2.23}$$

and the determinant of M^{-1} reduces to the especially convenient form

$$|M^{-1}| = \frac{(\Lambda_{11} g_{22} - \Lambda_{12} g_{22})^2}{\Lambda_{11} C_{22} - g_{21}^2} \tag{2.24}$$

and the denomination of (2.24) is $|C|$.

III. Single Parameter Settings

Poisson.

The Poisson random variable X has density

$$f(x;\lambda) = e^{-\lambda}\lambda^x/x! \qquad \text{for} \quad x = 0,1,... \tag{3.1}$$

and the derivative of the log likelihood is

$$S = \frac{X}{\lambda} - 1 \tag{3.2}$$

It is well-known that

$$\Lambda = E(S^2) = \frac{1}{\lambda} \tag{3.3}$$

and the sample mean is the minimum variance unbiased estimator for all sample sizes. It is also the maximum likelihood estimate and the method of moments estimate.

Let us look further solely for academic purposes. Since the Poisson has the property that its mean is equal to its variance it follows that the sample variance s^2 is also a "moment" estimate of λ. Moreover, the idea

can be extended. There are many other statistics that can be equated to their expected values and the resulting equations solved uniquely for λ. One other will be considered here, namely, the averages of reciprocals of the $\{1 + x_i\}_1^n$. The one is added as a convenient device for avoiding division by zero. The result will be called the harmonic mean estimator.

Since s^2 is directly an estimate of λ, its asymptotic efficiency is quickly and easily expressed. Using (2.22), (3.3), and (B.5) we have

$$\mathrm{Eff}(s^2) = \frac{\lambda}{\lambda + 2\lambda^2} \tag{3.4}$$

The harmonic mean estimator is characterized by the equation (see (B.7))

$$\bar{y} - \frac{1}{\lambda}(1 - e^{-\lambda}) = 0 \tag{3.5}$$

where

$$\bar{y} = \frac{1}{n}\sum_1^n \frac{1}{1 + x_i} \tag{3.6}$$

Equation (3.5) is in the form $g = 0$ (i.e. (2.1)) and satisfies the assumptions. The resulting estimate, λ^*, can be computed using the iteration function that falls naturally out of (3.5), namely,

$$\lambda_{r+1} = \frac{1}{\bar{y}}(1 - e^{-\lambda_r}) \tag{3.7}$$

and $\lambda_r \to \lambda^*$ for all initializations $\lambda_0 > 0$, but convergence can be quite slow if a poor λ_0 is chosen. Then asymptotic variance C (see (2.6)) of (3.5) is the variance of $(1 + X)^{-1}$ which can be computed from (B.7) and (B.8). The quantity A of (2.2) is obtained by differentiating (3.5) with respect to λ,

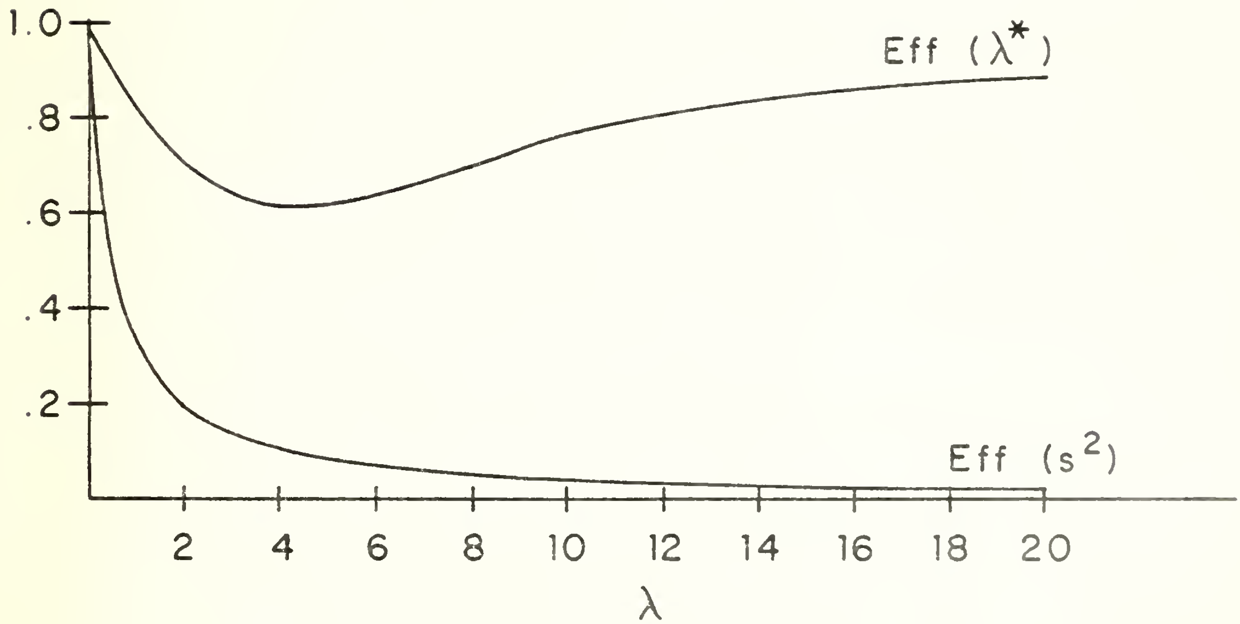
$$A = \frac{1}{\lambda^2}(1 - e^{-\lambda} - \lambda e^{-\lambda}) . \tag{3.8}$$

Using (2.22), (2.7) we have $\mathrm{Eff}(\lambda^*) = A^2\lambda/C$, or more explicitly

$$\mathrm{Eff}(\lambda^*) = \frac{1}{\lambda^3}(1 - e^{-\lambda} - \lambda e^{-\lambda})^2/\mathrm{Var}(\tfrac{1}{1+X}) \tag{3.9}$$

The efficiences of the two estimators are compared in Figure 3.1. Of course there is no point in using any estimator other than $\hat{\lambda} = \bar{x}$, but the graph suggests that the harmonic mean may be profitably used in other problems in which a choice must be exercised.

Figure 3.1
Asymptotic Efficiency
(Poisson)



Just for fun, let us compare the values of the three estimators $\bar{x}$, s^2, λ^* when applied to some famous data. First, the number of deaths due to mule kick in 200 Prussian army corps years. The frequency counts are 109, 65, 22, 3, 1 for variable values 0 thru 4 [7, p. 109]. The estimators are

$$\bar{x} = 0.61 , \qquad s^2 = .6109 , \qquad \lambda^* = .6093$$

and agreement is rather good. Second (Rutherford Chadwick data), the number of radioactive disintegration in 2608 time intervals of 7.5 seconds each [4, p. 150]. This time we have

$$\bar{x} = 3.867 , \qquad s^2 = 3.633 , \qquad \lambda^* = 3.886 .$$

The sample size is much larger but the value of λ is in a range of lower efficiency for s^2. Also, the radioactive decay is more properly modeled with a pure death process and this may help explain the smaller value of s^2.

One final comment. The convergence of the iteration function (3.7) has importance in finding the maximum likelihood estimate for λ from a Poisson population in which the zero values have been truncated. That is, a trial that produces no counts does not come to the experimenter's attention [10, p. 3-13]. The density is

$$f(x,\lambda) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \frac{\lambda^x}{x!} , \qquad x = 1,2, \ldots$$

and the maximum likelihood equation is

$$\bar{x} = \frac{\lambda}{1 - e^{-\lambda}}$$

which, as a function of λ, is the same as (3.5).

Symmetric Beta

A Beta random variable with equal parameter values has density

$$f(x,\alpha) = \frac{\Gamma(2\alpha)}{[\Gamma(\alpha)]^2} x^\alpha (1-x)^\alpha \qquad \text{for} \quad 0 \le x \le 1,\ 0 < \alpha \tag{3.10}$$

Using the psi function [1],

$$\psi(\alpha) = \frac{d \ln \Gamma(\alpha)}{d\alpha} \tag{3.11}$$

one can write the derivative of the log density as

$$S_\alpha = 2\psi(2\alpha) - 2\psi(\alpha) + \ln(x(1-x)) \tag{3.12}$$

and

$$\Lambda = 2\psi'(\alpha) - 4\psi'(2\alpha) \tag{3.13}$$

by (2.9). Let $\overline{\ln x} = \frac{1}{n} \sum_{i=1}^n \ln x_i$ and express the maximum likelihood equation as

$$\psi(\alpha) - \psi(2\alpha) = \frac{1}{2} \{\overline{\ln x} + \overline{\ln(1-x)}\} \tag{3.14}$$

so $\hat{\alpha}$, the maximum likelihood estimate, is a function of the geometric mean of x and $1-x$. Also, it is difficult to find.

Let us turn to the method of moments. Because of symmetry, see (E.11), (E.12),

$$\mu = \frac{1}{2}, \qquad \sigma^2 = \frac{1}{4(2\alpha + 1)}$$

Thus $\bar{x}$ cannot be used and we turn to the second moment. The form $g = 0$ becomes

$$s^2 - \frac{1}{4(2\alpha + 1)} = 0 \tag{3.15}$$

and the estimator can be found explicitly

$$\tilde{\alpha} = \frac{1}{8s^2} - \frac{1}{2} \tag{3.16}$$

and using (2.2), (2.7) produces

$$n\,\mathrm{Var}(\tilde{\alpha}) = 4(2\alpha + 1)^4\,\mathrm{Var}(s^2) \tag{3.17}$$

The right side of (3.17) requires that (E.2) be used with (A.2). The result is not compact and is produced only by computer program.

A harmonic mean option is available since

$$E(\frac{1}{X}) = \frac{2\alpha - 1}{\alpha - 1} = E(\frac{1}{1-X}) \tag{3.18}$$

(see (E.13)) for $\alpha > 1$, and the variances are finite if $\alpha > 2$. Easy calculations, exploiting (E.14) and (E.15), show that both y and z, where

$$y = \frac{1}{n}\sum_1^n \frac{1}{x_i}\ , \qquad z = \frac{1}{n}\sum_1^n \frac{1}{1-x_i}\ , \tag{3.19}$$

should be used instead of either one alone in order to reduce variability. We select the form $g = 0$ to be

$$(\alpha - 1)(y + z) - (4\alpha - 2) = 0 \tag{3.20}$$

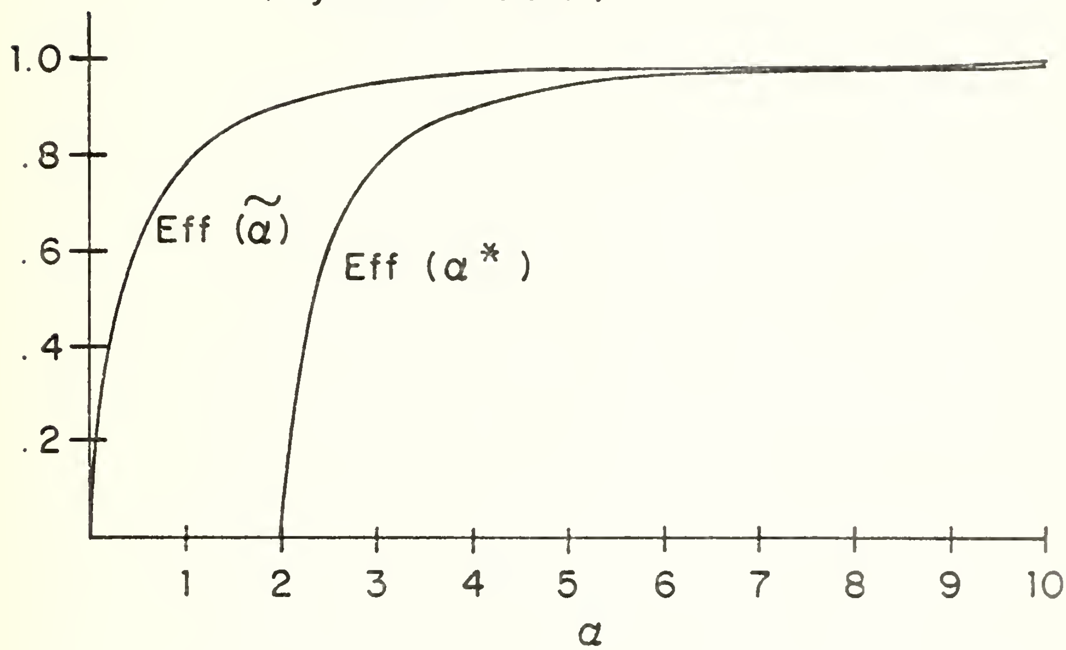
and solve explicitly

$$\alpha^* = \frac{y + z - 2}{y + z - 4} \tag{3.21}$$

Application of (2.2), (2.7), (E.14), and (E.15) produces

$$n\,\mathrm{Var}(\alpha^*) = \frac{(2\alpha - 1)(\alpha - 1)^2}{\alpha - 2} \qquad \text{for } \alpha > 2 \tag{3.22}$$

Figure 3.2
Asymptotic Efficiency
(Symmetric Beta)



The efficiencies of these two estimates are compared in Figure 3.2. The second moment estimator, $\tilde{\alpha}$, is uniformly better than the harmonic mean estimator, α^*, and this result needs interpretation in the light of the success in the Poisson case. This time the averages of X^{-1} and $(1-X)^{-1}$ were used instead of $(1+X)^{-1}$, the latter being difficult to manage with the beta distribution, and the parameter must be at least two in order for the variance to exist. Also the positive sample space is $0 < x < 1$, which entails large and variable values for the $\{x_i^{-1}\}$, and this may explain the lack of success. The estimator α^* may still have some uses because (unlike $\tilde{\alpha}$) we have a simple explicit expression for its variance (3.22) and its efficiency is competitive for α more than (say) four or five.

IV. Two Parameter Settings

Gamma

The gamma random variable X has density

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \tag{4.1}$$

and the partial derivatives of its logarithm are

$$\begin{aligned} S_\beta &= \frac{x}{\beta^2} - \frac{\alpha}{\beta} \\ S_\alpha &= \ln x - \ln \beta - \psi(\alpha) \end{aligned} \tag{4.2}$$

where $\psi(\alpha)$ is the psi function (derivative of log gamma). The maximum likelihood estimators $(\hat{\alpha},\hat{\beta})$ is found from (4.2) by replacing x with $\bar{x}$ and $\ln x$ with $\overline{\ln x} = \frac{1}{n}\sum_1^n \ln x_i$. Thus it fits into our generalized moment scheme utilizing the arithmetic and geometric means. To solve the system, one sets $\beta = \bar{x}/\alpha$ into the second member of (4.2) and search for the root of $\ln \alpha - \psi(\alpha) = \ln \bar{x} - \overline{\ln x}$, which requires a psi function capability [1,3]. This is not difficult with a large computer, but may be a challenge for a small one, e.g. a hand held calculator. Viable alternatives are available using the ordinary method of moments and a generalization that exploits the harmonic mean.

The information matrix (2.9) is

$$\Lambda = \begin{pmatrix} \alpha/\beta^2 & 1/\beta \\ 1/\beta & \psi'(\alpha) \end{pmatrix} \tag{4.3}$$

whose inverse is

$$\Lambda^{-1} = \frac{1}{\alpha\psi'(\alpha)-1} \begin{Bmatrix} \beta^2\psi'(\alpha) & -\beta \\ -\beta & \alpha \end{Bmatrix} \tag{4.4}$$

where primes denote derivative.

The ordinary method of moments equates $\bar{x}$ and s^2 to $\alpha\beta$ and $\alpha\beta^2$, respectively, and the resulting estimator is

$$\tilde{\beta} = s^2/\bar{x} , \qquad \tilde{\alpha} = \bar{x}^2/s^2 \tag{4.5}$$

Using $\theta_1 = \beta$ and $\theta_2 = \alpha$ in order to conform with (4.2), one can apply (2.2) to obtain

$$\tilde{A} = \begin{pmatrix} \alpha & \beta \\ 2\alpha\beta & \beta^2 \end{pmatrix} \tag{4.6}$$

Note: Although the method of moments shares an equation with maximum likelihood $(\bar{x}/\beta^2 - \alpha/\beta = 0)$, there is no advantage to using this, through (2.18), in this case.

The matrix C is obtained from (C.5), (C.6), and (C.7). Then (2.7) can be applied to get

$$\tilde{M} = 2(\alpha+1) \begin{Bmatrix} \frac{\beta^2}{\alpha}\,\frac{2\alpha+3}{2(\alpha+1)} & -\beta \\ -\beta & \alpha \end{Bmatrix} \tag{4.7}$$

and

$$|\tilde{M}| = 2(\alpha+1)\beta^2 \tag{4.8}$$

Turning to another choice, let us recognize that the ordinary method of moments uses moments of order one and two, while maximum likelihood uses moments of order one and zero. Heuristically one might find advantage

in moments of order one and minus one. Consider for the system (2.2) (see (C.2) and (C.8)),

$$\begin{aligned} \bar{x} - \alpha\beta &= 0 \\ y - \frac{1}{\beta(\alpha-1)} &= 0 \end{aligned} \tag{4.9}$$

where $y = \frac{1}{n}\sum_1^n 1/x_i$ and the consequent estimators

$$\beta^* = \bar{x} - \frac{1}{y} \qquad\qquad \alpha^* = \frac{\bar{x}y}{\bar{x}y - 1} \tag{4.10}$$

which satisfies $\beta^* > 0$ and $\alpha^* > 0$ since the arithmetic mean is larger than the harmonic mean. Proceeding as before we obtain, for $\alpha > 2$,

$$A^* = \begin{Bmatrix} -\alpha & -\beta \\ \dfrac{1}{\beta^2(\alpha-1)} & \dfrac{1}{\beta(\alpha-1)^2} \end{Bmatrix} \tag{4.11}$$

and

$$M^* = \frac{2(\alpha-1)^2}{\alpha - 2} \begin{Bmatrix} \beta^2 \dfrac{2\alpha - 3}{2(\alpha-1)^2} & -\beta \\ -\beta & \alpha \end{Bmatrix} \tag{4.12}$$

and

$$|M^*| = 2\beta^2 \frac{(\alpha-1)^2}{\alpha - 2} \qquad\qquad \text{for} \quad \alpha > 2 \tag{4.13}$$

The asymptotic efficiencies of the estimators (4.5) and (4.10) are computed from (2.22) using

$$|\Lambda| = \frac{1}{\beta^2}\,(\alpha\psi'(\alpha) - 1) \tag{4.14}$$

These efficiencies do not depend on the scale parameter, β, and can be

plotted as functions of α, as is done in Figure 4.1. Also, the former appears in [15]. The alternative (α^*,β^*) catches up to $(\tilde{\alpha},\tilde{\beta})$ at $\alpha = 3$ and remains better thereafter. The relative efficiency of the latter with respect to the former is (see (4.13) and (4.8))

$$\frac{|M^*|}{|\tilde{M}|} = \frac{(\alpha-1)^2}{(\alpha-2)(\alpha+1)} \tag{4.15}$$

It is less than 100% for all $\alpha > 3$ and falls below 90% for $4 < \alpha < 7$.

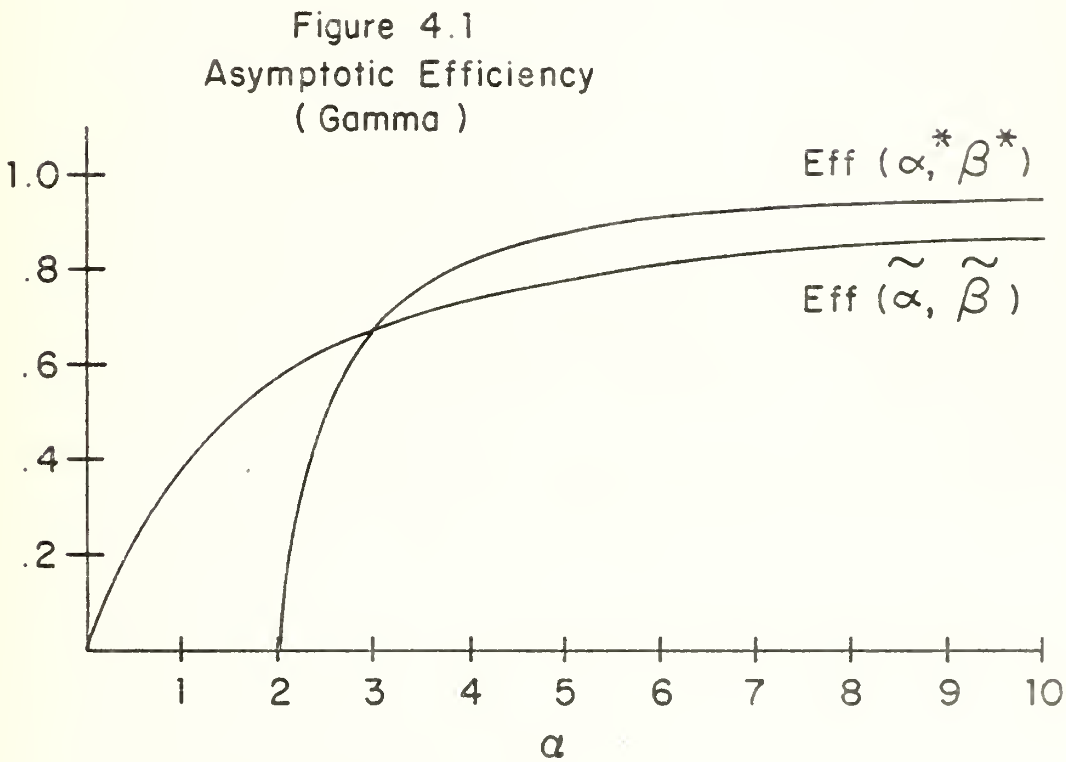


FIGURE 4.1. Asymptotic Efficiency (Gamma)

As a numerical example, the three estimators were applied to 135 observations on the service time of the check-in queue at a major hotel. The gamma distribution fits quite acceptably according to chi square criteria. The results are:

$$\hat{\beta} = .303 \qquad \tilde{\beta} = .305 \qquad \beta^{*} = .298$$

$$\hat{\alpha} = 6.33 \qquad \tilde{\alpha} = 6.31 \qquad \alpha^{*} = 6.45$$

Counterexample: This opportunity is taken to show that the representation (2.18), which is the main theorem of [12] , is false when one drops the assumption that the subsystems $\mu = 0$ (2.11) are likelihood equations. We use the system

$$y - \frac{1}{\beta(\alpha-1)} = 0$$

$$s^2 - \alpha\beta^2 = 0$$

which has an equation in common with the method of moments (4.5), but this common equation is not part of the maximum likelihood system. This system has explicit solutions

$$\tilde{\beta}^{*} = \frac{1}{2}\left\{-\frac{1}{y} + \sqrt{\frac{1}{y^2} + 4s^2}\right\}, \qquad \tilde{\alpha}^{*} = 1 + \frac{1}{y\tilde{\beta}^{*}}$$

One readily develops, using (4.11) and (4.6)

$$\tilde{A}^{*} = \begin{Bmatrix} \dfrac{1}{\beta^2(\alpha-1)} & \dfrac{1}{\beta(\alpha-1)^2} \\ 2\alpha\beta & \beta^2 \end{Bmatrix}$$

and from (C.7), (C.9), (C.11)

$$\tilde{C}^* = \begin{pmatrix} \dfrac{1}{\beta^2(\alpha-1)^2\ (\alpha-2)} & 0 \\ 0 & 2\alpha\beta^4(\alpha+3) \end{pmatrix}$$

Using $M^{-1} = A'C^{-1}A$, we calculate

$$\tilde{M}^{*-1} = \begin{pmatrix} \dfrac{1}{\beta^2(\alpha-1)^2(\alpha-2)} + \dfrac{2\alpha}{\beta^2(\alpha+3)} & \dfrac{\alpha-2}{\beta(\alpha-1)} + \dfrac{1}{\beta(\alpha+3)} \\ \dfrac{\alpha-2}{\beta(\alpha-1)} + \dfrac{1}{\beta(\alpha+3)} & \dfrac{\alpha-2}{(\alpha-1)^2} + \dfrac{1}{2\alpha(\alpha+3)} \end{pmatrix}$$

and this, clearly, has no submatrix in common with (upon inverting (4.7))

$$\tilde{M}^{-1} = \frac{1}{\beta^2} \begin{pmatrix} \alpha & \beta \\ \beta & \dfrac{\beta^2}{\alpha}\ \dfrac{2\alpha+3}{2(\alpha+1)} \end{pmatrix}$$

Negative Binomial

The negative binomial density is parametrized

$$f(x;r,p) = \frac{\Gamma(r+x)}{x!\ \Gamma(r)}\, q^x p^r \qquad \text{for} \quad x = 0,1,2,\ldots \tag{4.16}$$

$$0 < r,\ 0 < p < 1,\ p+q = 1$$

and the partial derivatives of its logarithm are

$$\begin{aligned} S_p &= r/p - x/q \\ S_r &= \psi(r+x) - \psi(r) + \ln p \end{aligned} \tag{4.17}$$

Using the basic recursive formula for the psi function [1],

$$\psi(r+x) - \psi(r) = \sum_{j=1}^{x} \frac{1}{r - 1 + j}$$

one may express the system of maximum likelihood equations as

$$\bar{x} - rq/p = 0$$

$$\operatorname*{ave}_{i=1,\ldots,n} \sum_{j=1}^{x_i} \frac{1}{r - 1 + j} + \ln p = 0 \tag{4.18}$$

Solving (4.18) is quite difficult to manage without a large computer. Appendix F contains an APL program (PSIB) which accomplishes it.

The information matrix can be developed using (2.9). The result is, using $\theta_1 = p$ and $\theta_2 = r$,

$$\Lambda = \begin{Bmatrix} r/qp^2 & -1/p \\ -1/p & \Lambda_{22} \end{Bmatrix} \tag{4.19}$$

where

$$\Lambda_{22} = \psi'(r) - E(\psi'(r+X)) \tag{4.20}$$

The properties of Λ_{22} are developed in Appendix D (see (D.22) and (D.26)), along with the series representation of the determinant (D.25),

$$|\Lambda| = \frac{1}{p^2} \sum_{n=1}^{\infty} \frac{q^n}{n + 1} \frac{r!\, n!}{(n+r)!} \tag{4.21}$$

which converges rapidly for most values of p. For purposes of computation, one may as well use (4.21) in (4.19) and solve for Λ_{22}, thus

$$\Lambda_{22} = \frac{1}{rq} \{1 + p^2|\Lambda|\} \tag{4.23}$$

To estimate by the ordinary method of moments, we use (D.5) and form the system (2.1) to be

$$
\begin{aligned}
p\bar{x} - rq &= 0 \\
p^2 s^2 - rq &= 0
\end{aligned}
\qquad (4.24)
$$

The explicit solution is readily seen to be

$$\tilde{p} = \frac{\bar{x}}{s^2}, \qquad \tilde{r} = \frac{\bar{x}^2}{s^2 - \bar{x}} \qquad (4.25)$$

and it cannot be guaranteed that $\tilde{p} < 1$ and $\tilde{r} > 0$. Differentiating (4.24) and taking expectations yields the matrix

$$\tilde{A} = \begin{pmatrix} r/q & -q \\ \dfrac{r(1+q)}{p} & -q \end{pmatrix} \qquad (4.26)$$

Combining (2.6), (4.24) and (D.6) produces

$$\tilde{C} = rq \begin{pmatrix} 1 & 1+q \\ 1+q & 1+2(r+2)q+q^2 \end{pmatrix} \qquad (4.27)$$

and finally the formula (2.7)

$$\tilde{M} = \frac{2(r+1)}{rq^2} \begin{pmatrix} p^2q^2\left(1 + \dfrac{q}{2(r+1)}\right) & pq \\ pq & r^2 \end{pmatrix} \qquad (4.28)$$

and

$$|\tilde{M}| = 2(r+1)p^2/q \qquad (4.29)$$

Let us turn to the question of using a different moment in the second equation. Since the negative binomial has positive probability mass at zero, we use averages of the $(1+x_i)^{-1}$ as was done when the Poisson case was considered. Perhaps that level of success can be matched.

Letting $y = \frac{1}{n}\sum_1^n 1/(1+x_i)$, use for the system (2.1)

$$\begin{aligned} p\bar{x} - rq &= 0 \\ (r-1)qy - p + p^r &= 0 \end{aligned} \tag{4.30}$$

because of (D.7). The system (4.30) cannot be solved explicitly, but it can be managed with a hand held calculator. Use the first member to obtain r as a function of p and its derivative

$$r = \frac{\bar{x}p}{q}, \qquad \frac{dr}{dp} = \frac{\bar{x}}{q^2} \tag{4.31}$$

and substitute into the second member of (4.30) to obtain a function f of the form

$$\begin{aligned} f(p) &= p^r + (r-1)_{qy} - p \\ &= p^r - y + p(y - 1 + \bar{x}y) \end{aligned} \tag{4.32}$$

and having derivatives

$$\frac{df}{dp} = (y - 1 + \bar{x}y) + \frac{p^r\bar{x}(q + \ln p)}{q^2}$$

The solution of $f(p) = 0$ can be obtained by Newton's method, always remembering to update r as well as p. The initialization $p = .5$ and $r = \bar{x}$ appears to be satisfactory, but normally convergence would be faster if the moment estimators (4.25) are used.

The resulting estimator will be denoted by p^*, r^*. From (4.32) it is seen that

$$f(0) = -y < 0 , \qquad\qquad f(1) = \bar{x}y > 0$$

so that $0 < p^* < 1$, and $r^* > 0$ follows from this by (4.31).

Let us turn to the development of the asymptotic covariance structure of p^*, r^*. Taking partial derivatives and the expectation of (4.32) to produce the coefficient matrix A of (2.2) yields

$$A^* = \begin{pmatrix} r/p & -q \\ rp^{r-1} - \dfrac{1-p^r}{q} & \dfrac{p-p^r}{r-1} + p^r \ln p \end{pmatrix} \qquad (4.33)$$

with the help of (D.7). Attention is drawn to the fact that

$$\lim_{r \to 1} \frac{p - p^r}{r - 1} = p - p \ln p \qquad (4.34)$$

The covariance matrix C of (2.6) is derived from (4.30) with the help of (D.5), (D.7), and (A.5). The result is

$$C^* = \begin{pmatrix} rq & rqp^r - p(1-p^r) \\ rqp^r - p(1-p^r) & (r-1)^2 q^2 \operatorname{Var}(\frac{1}{1+X}) \end{pmatrix} \qquad (4.34)$$

Let us draw attention to the fact the first equation of (4.17) is a multiple of the first equation of (4.32) rather than being identical. This is the reason that (4.33) and (4.35) are modifications of (2.17) and (2.16) rather than exact analogies. Thus the bookkeeping that follows must be done carefully.

The direct development of the matrix M^* from (2.7) with (4.34) and (4.35) as input is messy. Instead let us recognize that (4.30) shares an equation with the likelihood system (4.18) and use (2.18). Thus, with Λ given by (4.19), we have

$$M^{*-1} = \begin{Bmatrix} r/qp^2 & -1/p \\ -1/p & M^{22} \end{Bmatrix} \tag{4.36}$$

where M^{22} is obtained from (2.23) and $|C|$ from (2.16). Thus

$$M^{22} = \frac{1}{|C|}\left\{\frac{r}{qp^2}\, g_{22}^2 + \frac{2}{p}\, g_{21}g_{22} + \frac{C_{22}}{p^2}\right\}$$
$$|C| = \frac{r}{qp^2}\, C_{22} - g_{21}^2 \tag{4.37}$$

and (g_{21}, g_{22}) is the second row of A^* in (4.33) and C_{22} is taken from (4.35). Using (2.24) we obtain

$$|M^*|^{-1} = \left(\frac{r}{qp^2}\, g_{22} + \frac{1}{p}\, g_{21}\right)^2 \bigg/ |C|$$

and (4.36)

$$|C| = \frac{r}{p^2}\,(r-1)^2 q \operatorname{Var}\left(\frac{1}{1+X}\right) - g_{21}^2$$

The asymptotic efficiencies of $(\tilde{p},\tilde{r})$ and (p^*,r^*) appear in Tables 4.1 and 4.2, resp., for $p = .1(.1).9$ and $r = .5(.5)5,\ 6(1)19$ where the parentheses indicate the indices of advancement. The efficiency of $(\tilde{p},\tilde{r})$ is monotone increasing in r for each p. It is lower for the smaller values of p. The efficiency of (p^*,r^*) is high for the smaller values of r and decreases (generally). It is not monotone for $p = .1, .2$.

The relative efficiency of p^*, r^* with respect to $\tilde{p}, \tilde{r}$, i.e.,

$$\text{Rel eff} = |M^*|^{-1} \cdot |\tilde{M}|$$

appears in Table 4.3. Generally p^*, r^* is preferable for r less than or equal to (say) 2.5. In general $\tilde{p}, \tilde{r}$ is preferable elsewhere, but it does not matter much for small values of p.

The three estimation schemes were applied to the Cricket score data of Reep, Pollard and Benjamin [13], which provided the following comparisons.

	Cowdry	Barrington	Graveney
$\bar{x}$	1.692	2.095	1.570
s^2	4.343	4.939	4.474
$\overline{(1+x)^{-1}}$	.603	.538	.626
n	156.	116.	107.
$\hat{p}$	.329	.345	.317
$\hat{r}$	.831	1.014	.729
$\tilde{p}$	.390	.424	.351
$\tilde{r}$	1.080	1.543	.849
p^*	.371	.326	.389
r^*	1.000	1.012	1.000

TABLE 4.1. ASYMPTOTIC EFFICIENCY OF $\tilde{p}$, $\tilde{r}$

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.409	0.501	0.578	0.648	0.713	0.775	0.834	0.891	0.946
1.0	0.507	0.582	0.647	0.705	0.760	0.812	0.861	0.909	0.955
1.5	0.582	0.644	0.698	0.748	0.794	0.839	0.881	0.922	0.961
2.0	0.639	0.691	0.738	0.780	0.821	0.859	0.896	0.932	0.966
2.5	0.684	0.728	0.768	0.806	0.841	0.875	0.908	0.939	0.970
3.0	0.720	0.758	0.793	0.826	0.858	0.888	0.917	0.945	0.973
3.5	0.748	0.782	0.813	0.843	0.871	0.898	0.925	0.950	0.975
4.0	0.791	0.818	0.844	0.868	0.892	0.914	0.937	0.958	0.979
4.5	0.808	0.832	0.855	0.878	0.900	0.921	0.941	0.961	0.981
5.0	0.834	0.855	0.875	0.894	0.913	0.931	0.949	0.966	0.983
6.0	0.854	0.872	0.889	0.906	0.923	0.939	0.954	0.970	0.985
7.0	0.870	0.886	0.901	0.916	0.931	0.945	0.959	0.973	0.987
8.0	0.882	0.897	0.910	0.924	0.937	0.950	0.963	0.975	0.988
9.0	0.893	0.906	0.918	0.930	0.943	0.954	0.966	0.977	0.989
10.0	0.902	0.913	0.925	0.936	0.947	0.958	0.969	0.979	0.990
11.0	0.909	0.920	0.930	0.941	0.951	0.961	0.971	0.981	0.990
12.0	0.915	0.925	0.935	0.945	0.954	0.964	0.973	0.982	0.991
13.0	0.921	0.930	0.939	0.948	0.957	0.966	0.975	0.983	0.992
14.0	0.926	0.935	0.943	0.951	0.960	0.968	0.976	0.984	0.992
15.0	0.930	0.938	0.946	0.954	0.962	0.970	0.977	0.985	0.993
16.0	0.934	0.942	0.949	0.957	0.964	0.971	0.979	0.986	0.993
17.0	0.937	0.945	0.952	0.959	0.966	0.973	0.980	0.987	0.993
18.0	0.941	0.947	0.954	0.961	0.968	0.974	0.981	0.987	0.994
19.0	0.943	0.950	0.956	0.963	0.969	0.975	0.982	0.988	0.994

TABLE 4.2. ASYMPTOTIC EFFICIENCY OF p^*, r^*

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.995	0.999	1.000	1.000	0.999	0.999	1.000	1.000	1.000
1.0	0.926	0.965	0.980	0.987	0.992	0.995	0.997	0.998	0.999
1.5	0.825	0.898	0.933	0.953	0.967	0.977	0.985	0.991	0.996
2.0	0.740	0.828	0.877	0.910	0.934	0.953	0.968	0.981	0.991
2.5	0.685	0.767	0.822	0.864	0.898	0.925	0.948	0.968	0.985
3.0	0.663	0.721	0.775	0.821	0.861	0.896	0.927	0.954	0.978
3.5	0.667	0.692	0.736	0.782	0.826	0.867	0.905	0.940	0.971
4.0	0.723	0.678	0.687	0.720	0.764	0.812	0.862	0.910	0.956
4.5	0.759	0.687	0.675	0.697	0.738	0.787	0.840	0.895	0.948
5.0	0.821	0.725	0.674	0.669	0.696	0.742	0.801	0.865	0.933
6.0	0.863	0.772	0.694	0.660	0.667	0.705	0.764	0.837	0.917
7.0	0.891	0.813	0.725	0.666	0.651	0.676	0.732	0.810	0.901
8.0	0.909	0.846	0.760	0.683	0.646	0.655	0.705	0.784	0.886
9.0	0.922	0.870	0.792	0.706	0.649	0.641	0.681	0.761	0.871
10.0	0.932	0.888	0.819	0.732	0.659	0.632	0.661	0.739	0.856
11.0	0.940	0.901	0.842	0.758	0.674	0.630	0.645	0.719	0.841
12.0	0.946	0.912	0.860	0.782	0.692	0.631	0.632	0.700	0.827
13.0	0.951	0.921	0.875	0.803	0.711	0.637	0.623	0.683	0.814
14.0	0.955	0.928	0.886	0.822	0.731	0.646	0.617	0.668	0.800
15.0	0.958	0.934	0.896	0.838	0.751	0.658	0.613	0.655	0.788
16.0	0.961	0.939	0.905	0.851	0.769	0.671	0.612	0.643	0.775
17.0	0.964	0.943	0.912	0.863	0.786	0.685	0.614	0.632	0.763
18.0	0.966	0.947	0.918	0.873	0.801	0.700	0.617	0.623	0.752
19.0	0.968	0.950	0.923	0.881	0.815	0.715	0.622	0.616	0.741

TABLE 4.3. RELATIVE EFFICIENCY OF p^*, r^* WITH RESPECT TO $\tilde{p}, \tilde{r}$

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
.5	2.433	1.994	1.728	1.542	1.401	1.290	1.199	1.122	1.057
1.0	1.828	1.657	1.514	1.399	1.305	1.226	1.157	1.098	1.046
1.5	1.419	1.395	1.336	1.274	1.217	1.165	1.118	1.075	1.036
2.0	1.157	1.197	1.189	1.166	1.138	1.109	1.080	1.052	1.026
2.5	1.001	1.053	1.070	1.073	1.067	1.057	1.045	1.031	1.016
3.0	0.921	0.952	0.977	0.994	1.004	1.010	1.011	1.009	1.006
3.5	0.892	0.886	0.905	0.928	0.948	0.966	0.979	0.989	0.996
4.0	0.914	0.829	0.814	0.829	0.857	0.888	0.920	0.950	0.976
4.5	0.939	0.826	0.789	0.794	0.820	0.855	0.893	0.931	0.967
5.0	0.984	0.849	0.771	0.749	0.762	0.797	0.844	0.896	0.949
6.0	1.011	0.885	0.781	0.728	0.723	0.751	0.801	0.863	0.931
7.0	1.024	0.918	0.805	0.727	0.700	0.716	0.764	0.832	0.913
8.0	1.031	0.943	0.834	0.739	0.689	0.690	0.732	0.804	0.897
9.0	1.033	0.960	0.862	0.759	0.688	0.671	0.705	0.778	0.880
10.0	1.034	0.972	0.886	0.782	0.696	0.660	0.682	0.755	0.865
11.0	1.034	0.980	0.905	0.805	0.708	0.655	0.664	0.733	0.849
12.0	1.033	0.986	0.920	0.827	0.725	0.655	0.650	0.713	0.835
13.0	1.032	0.990	0.931	0.847	0.743	0.660	0.639	0.695	0.821
14.0	1.031	0.993	0.940	0.864	0.762	0.668	0.632	0.679	0.807
15.0	1.030	0.995	0.947	0.878	0.781	0.678	0.628	0.665	0.794
16.0	1.029	0.997	0.953	0.890	0.798	0.691	0.626	0.652	0.781
17.0	1.028	0.998	0.958	0.900	0.814	0.704	0.626	0.641	0.769
18.0	1.027	0.999	0.962	0.908	0.828	0.719	0.629	0.631	0.757
19.0	1.026	1.000	0.965	0.916	0.841	0.733	0.634	0.623	0.745

Beta

The beta density has the form

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \tag{4.37}$$

for $0 < x < 1,\ 0 < \alpha,\ 0 < \beta$

and the partial derivatives of its logarithm are

$$\begin{aligned} S_\alpha &= \psi(\alpha+\beta) - \psi(\alpha) + \ln x \\ S_\beta &= \psi(\alpha+\beta) - \psi(\beta) + \ln(1-x) \end{aligned} \tag{4.38}$$

The information matrix is, for $\theta_1 = \alpha,\ \theta_2 = \beta$

$$\Lambda = \begin{pmatrix} \psi'(\alpha) - \psi'(\alpha+\beta) & -\psi'(\alpha+\beta) \\ -\psi(\alpha+\beta) & \psi'(\beta) - \psi'(\alpha+\beta) \end{pmatrix} \tag{4.39}$$

The system of maximum likelihood equations

$$\begin{aligned} \overline{\ln x} &= \psi(\alpha) - \psi(\alpha+\beta) \\ \ln(1-x) &= \psi(\beta) - \psi(\alpha+\beta) \end{aligned} \tag{4.40}$$

uses the geometric means of x and $1-x$, and is difficult to solve. What other pairs of statistics might be substituted?

Clearly $\bar{x}$ and $\overline{(1-x)}$ cannot be paired since they are functionally related. The latter is merely $1-\bar{x}$. Let us first develop the ordinary method of moments. Using (E.4) and (E.5), choose the systems

$$\begin{aligned} (\alpha+\beta)\bar{x} - \alpha &= 0 \\ (\alpha+\beta)^2\,(\alpha+\beta+1)s^2 - \alpha\beta &= 0 \end{aligned} \tag{4.41}$$

and solve explicitly for

$$\tilde{\alpha} = \frac{s^2 - \bar{x}(1-\bar{x})}{1-\bar{x}} , \qquad \tilde{\beta} = \frac{s^2 - \bar{x}(1-\bar{x})}{\bar{x}} \tag{4.42}$$

It may occur that $\tilde{\alpha} < 0$, $\tilde{\beta} < 0$.

The coefficient matrix A of (2.2) is

$$\tilde{A} = \frac{1}{\alpha+\beta} \begin{pmatrix} -\beta & \alpha \\ -\beta(\frac{\alpha}{\alpha+\beta+1} + \beta) & -\alpha(\frac{\beta}{\alpha+\beta+1} + \alpha) \end{pmatrix} \tag{4.43}$$

$$|\tilde{A}| = \frac{-\alpha\beta}{\alpha + \beta + 1} \tag{4.44}$$

and C of (2.6) takes the form

$$\tilde{C} = \begin{pmatrix} \frac{\alpha\beta}{\alpha+\beta+1} & (\alpha+\beta)^3(\alpha+\beta+1)\ \mathrm{Cov}(\bar{x},s^2) \\ (\alpha+\beta)^3(\alpha+\beta+1)\ \mathrm{Cov}(\bar{x},s^2) & (\alpha+\beta)^4(\alpha+\beta+1)^2\ \mathrm{var}(s^2) \end{pmatrix} \tag{4.45}$$

and the use of (E.2) thru (A.1) and (A.2) does not appear to simplify in any useful way. Appendix F contains programs to compute $|M| = |C|/|A|^2$.

Let us turn to the pair of statistics

$$y = \frac{1}{n}\sum_1^n \frac{1}{x_i} \qquad z = \frac{1}{n}\sum_1^n \frac{1}{1-x_i} \tag{4.46}$$

which are not functionally related. Using (E.8) we may choose the system

$$\begin{aligned} (\alpha-1)y - (\alpha+\beta-1) &= 0 \\ (\beta-1)z - (\alpha+\beta-1) &= 0 \end{aligned} \tag{4.47}$$

for $\alpha > 1$, $\beta > 1$. The solution is

$$\alpha^* = \frac{y(z-1)}{yz - y - z} \qquad \beta^* = \frac{z(y-1)}{yz - y - z} \tag{4.48}$$

Clearly $y > 1$ and $z > 1$ but the denominators are not necessarily positive since $zy - y - z + 1 = (y-1)(z-1)$ can be less than one.

For the system (4.47) the coefficient matrix becomes

$$A^* = \begin{pmatrix} \frac{\beta}{\alpha-1} & -1 \\ -1 & \frac{\alpha}{\beta-1} \end{pmatrix} \tag{4.49}$$

and with the help of (E.9) and (E.10) one can calculate

$$C^* = (\alpha+\beta-1) \begin{pmatrix} \frac{\beta}{\alpha-2} & -1 \\ -1 & \frac{\alpha}{\beta-2} \end{pmatrix} \tag{4.50}$$

for $\alpha > 2$, $\beta > 2$. Use of (4.49) and (4.50) in (2.7) does not produce a convenient expression for M^*. However its determinant is easily managed.

For fun, let us also try a system based on moments of order one and minus one. (This would seem to straddle the two geometric means appearing in maximum likelihood.) We are lead to

$$\begin{aligned} (\alpha+\beta)\bar{x} - \alpha &= 0 \\ (\alpha-1)y - (\alpha+\beta-1) &= 0 \end{aligned} \tag{4.51}$$

and the resulting estimate

$$\tilde{\alpha}^* = \frac{(y-1)\bar{x}}{y\bar{x}-1} \qquad \tilde{\beta}^* = \frac{(y-1)(1-\bar{x})}{y\bar{x}-1} \tag{4.52}$$

and this satisfies $\tilde{\alpha}^* > 0$, $\tilde{\beta}^* > 0$. (But do not forget that use of (E.6) in (4.51) requires $\alpha > 1$.) Proceeding in the usual way, we calculate

$$\tilde{A}^* = \begin{pmatrix} -\dfrac{\beta}{\alpha + \beta} & \dfrac{\alpha}{\alpha + \beta} \\ \dfrac{\beta}{\alpha - 1} & -1 \end{pmatrix} \tag{4.53}$$

$$\tilde{C}^* = \begin{pmatrix} \dfrac{\alpha\beta}{\alpha + \beta + 1} & -\beta \\ -\beta & \dfrac{\beta(\alpha+\beta+1)}{\alpha - 2} \end{pmatrix} \qquad 4.54)$$

Again $\tilde{M}^*$ does not have a convenient form. The determinates of (4.52) and (4.53) are expressed

$$|A| = \frac{-\beta}{(\alpha+\beta)(\alpha-1)}$$
$$|C| = \frac{2\beta^2}{\alpha - 2} \tag{4.55}$$

for $\alpha > 2$.

The efficiencies of (4.42), (4.48), and (4.52) are compared in the tables that follow. Table 4.4 contains the efficiency of the ordinary moment estimator. Efficiency is high if α is not too far from β and both are at least two. Elsewhere they are low, but still the choice because all the numbers in Table 4.4 are better than their competitors in Tables 4.5 and 4.6. The pair (α^*, β^*) is generally better than $(\tilde{\alpha}^*, \tilde{\beta}^*)$ but not uniformly so. It matters little since $(\tilde{\alpha}, \tilde{\beta})$ is the "hands down" winner. This result parallels what was learned in the symmetric beta case. The variable X^{-1} is unstable in this population.

TABLE 4.4. Eff($\tilde{\alpha}$,$\tilde{\beta}$) BETA POPULATION

	.5	1.0	1.5	2.0	2.5	3.0	3.5
.5	0.493	0.554	0.537	0.507	0.477	0.451	0.429
1.0	0.554	0.713	0.747	0.739	0.719	0.696	0.674
1.5	0.537	0.747	0.820	0.839	0.835	0.822	0.806
2.0	0.507	0.739	0.839	0.878	0.889	0.887	0.878
2.5	0.477	0.719	0.835	0.889	0.912	0.919	0.918
3.0	0.451	0.696	0.822	0.887	0.919	0.934	0.938
3.5	0.429	0.674	0.806	0.878	0.918	0.938	0.948
4.0	0.411	0.653	0.790	0.867	0.912	0.938	0.952
4.5	0.395	0.635	0.774	0.855	0.904	0.933	0.951
5.0	0.382	0.618	0.758	0.843	0.895	0.928	0.948
5.5	0.370	0.604	0.744	0.831	0.886	0.921	0.944
6.0	0.360	0.591	0.731	0.819	0.876	0.914	0.938
6.5	0.351	0.579	0.719	0.809	0.867	0.906	0.933
7.0	0.344	0.568	0.709	0.799	0.858	0.899	0.927

	4.0	4.5	5.0	5.5	6.0	6.5	7.0
.5	0.411	0.395	0.382	0.370	0.360	0.351	0.344
1.0	0.653	0.635	0.618	0.604	0.591	0.579	0.568
1.5	0.790	0.774	0.758	0.744	0.731	0.719	0.709
2.0	0.867	0.855	0.843	0.831	0.819	0.809	0.799
2.5	0.912	0.904	0.895	0.886	0.876	0.867	0.858
3.0	0.938	0.933	0.928	0.921	0.914	0.906	0.899
3.5	0.952	0.951	0.948	0.944	0.938	0.933	0.927
4.0	0.959	0.961	0.961	0.958	0.955	0.951	0.946
4.5	0.961	0.966	0.968	0.968	0.966	0.963	0.960
5.0	0.961	0.968	0.972	0.973	0.973	0.972	0.969
5.5	0.958	0.968	0.973	0.976	0.977	0.977	0.976
6.0	0.955	0.966	0.973	0.977	0.980	0.980	0.980
6.5	0.951	0.963	0.972	0.977	0.980	0.982	0.983
7.0	0.946	0.960	0.969	0.976	0.980	0.983	0.984

TABLE 4.5. $Eff(\alpha^*, \beta^*)$ BETA POPULATION

	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
2.5	0.352	0.462	0.506	0.526	0.535	0.539	0.540	0.540	0.540	0.539
3.0	0.462	0.613	0.676	0.706	0.721	0.729	0.732	0.734	0.734	0.733
3.5	0.506	0.676	0.749	0.786	0.805	0.815	0.821	0.824	0.825	0.826
4.0	0.526	0.706	0.786	0.826	0.848	0.861	0.869	0.873	0.875	0.876
4.5	0.535	0.721	0.805	0.848	0.873	0.887	0.896	0.902	0.905	0.907
5.0	0.539	0.729	0.815	0.861	0.887	0.903	0.913	0.919	0.923	0.926
5.5	0.540	0.732	0.821	0.869	0.896	0.913	0.924	0.931	0.936	0.939
6.0	0.540	0.734	0.824	0.873	0.902	0.919	0.931	0.939	0.944	0.947
6.6	0.540	0.734	0.825	0.875	0.905	0.923	0.936	0.944	0.950	0.954
7.0	0.539	0.733	0.826	0.876	0.907	0.926	0.939	0.947	0.954	0.958

TABLE 4.6. Eff($\tilde{\alpha}^*, \tilde{\beta}^*$) BETA POPULATION

	.5	1.0	1.5	2.0	2.5	3.0	3.5
2.5	0.119	0.202	0.258	0.297	0.326	0.347	0.364
3.0	0.163	0.276	0.353	0.407	0.447	0.477	0.500
3.5	0.184	0.313	0.400	0.461	0.506	0.540	0.566
4.0	0.196	0.333	0.426	0.492	0.540	0.576	0.605
4.5	0.204	0.346	0.443	0.511	0.561	0.599	0.629
5.0	0.209	0.355	0.454	0.524	0.576	0.615	0.645
5.5	0.212	0.361	0.462	0.534	0.586	0.626	0.657
6.0	0.215	0.366	0.468	0.541	0.594	0.634	0.666
6.5	0.217	0.369	0.473	0.546	0.600	0.641	0.673
7.0	0.218	0.372	0.476	0.550	0.604	0.645	0.678

	4.0	4.5	5.0	5.5	6.0	6.5	7.0
2.5	0.378	0.389	0.398	0.406	0.412	0.418	0.423
3.0	0.519	0.534	0.546	0.557	0.566	0.574	0.581
3.5	0.588	0.605	0.619	0.632	0.642	0.651	0.659
4.0	0.627	0.646	0.662	0.675	0.686	0.695	0.704
4.5	0.653	0.672	0.688	0.702	0.714	0.724	0.733
5.0	0.670	0.690	0.707	0.721	0.733	0.743	0.752
5.5	0.682	0.703	0.720	0.734	0.746	0.757	0.766
6.0	0.691	0.712	0.730	0.744	0.757	0.768	0.777
6.5	0.698	0.719	0.737	0.752	0.765	0.776	0.785
7.0	0.704	0.725	0.743	0.758	0.771	0.782	0.792

APPENDIX A

General Variance and Covariance Formulae Connecting the Mean, Variance, Harmonic Mean

Let $X_1,\ldots,X_n$ be a random sample of size n and denote the sample mean with

$$\bar{X} = \frac{1}{n}\sum_1^n X_i ,$$

the sample variance with

$$s^2 = \frac{1}{n-1}\sum_1^n (X_i-\bar{X})^2 ,$$

the inverse of the harmonic mean with

$$Y = \frac{1}{n}\sum_1^n \frac{1}{X_i} ,$$

and the inverse of the shifted harmonic mean with

$$Y' = \frac{1}{n}\sum_1^n \frac{1}{1+X_i} .$$

It is well known that $E(\bar{X}) = \mu$, $E(s^2) = \sigma^2$ when the population mean, μ, and variance, σ^2, exist. Let $m_r = E(X^r)$ and $\mu_r = E(X-\mu)^r$. The following relationships are needed.

$$\begin{aligned} \mathrm{Cov}(\bar{X},s^2) &= \frac{1}{n}\{m_3 - 3m_2m_1 + 2m_1^3\} \\ &= \frac{1}{n}\mu_3 \qquad (A.1) \end{aligned}$$

$$\mathrm{Var}(s^2) = \frac{1}{n}\{m_4 - 4m_3m_1 + 3m_2^2 - 4\sigma^4 + \frac{2}{n-1}\sigma^4\} \tag{A.2}$$

$$= \frac{1}{n}\{\mu_4 - \sigma^4 + \frac{2}{n-1}\sigma^4\}$$

$$\mathrm{Var}(\frac{1}{n}\sum_1^n (X_i-\bar{X})^2) = \frac{\mu_4 - \mu_2^2}{n} - 2\,\frac{(\mu_4-2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3} \tag{A.3}$$

$$\mathrm{Cov}(\bar{X},Y) = \frac{1}{n}\{1 - E(X)\ E(1/X)\} < 0 \tag{A.4}$$

$$\mathrm{Cov}(\bar{X},Y') = \frac{1}{n}\{1 - [1 + E(X)]\ E(\frac{1}{1+X})\} < 0 \tag{A.5}$$

$$\mathrm{Cov}(s^2,Y) = -\frac{1}{n}\{m_2E(1/X) + m_1 - 2m_1^2E(1/X)\} \tag{A.6}$$

$$= -\frac{1}{n}\{\mu_2m_{-1} - \mu^2m_{-1} + \mu\}$$

$$\mathrm{Cov}(s^2,Y') = -\frac{1}{n}\{[\mu_2 - (1+\mu)^2]\ E(\frac{1}{1+X}) + 1 + \mu\} \tag{A.7}$$

The proofs of (A.1) to (A.7) follow. It is convenient to record the relationships

$$m_2 = \mu_2 + \mu^2$$

$$m_3 = \mu_3 + 3\mu_2\mu + \mu^3 \tag{A.8}$$

$$m_4 = \mu_4 + 4\mu_3\mu + 6\mu_2\mu^2 + \mu^4$$

To enhance the readability, the symbols E, V, C will be used to denote expectation, variance, and covariance (resp.). Parentheses and subscripts will be used sparingly.

Proof of (A.4). Consider

$$E\bar{X}Y = \frac{1}{n^2} E\Sigma X_i \Sigma(1/X_j)$$

$$= \frac{1}{n^2} \{n + n(n-1)\ EXE(1/X)\}$$

from which we subtract $EXE(1/X)$ to produce (A.4). The fact that (A.4) is negative follows from the presumption that $X > 0$ and the consequence that the harmonic mean is less than the arithmetic mean. Thus (ave X_i) $\times$ (ave $\frac{1}{X_i}$) unless (all x_i = constant) and apply the law of large numbers.

Proof of (A.5). Follows from the fact that $C(1 + \bar{X})Y' = C\bar{X}Y'$ and the application of (A.4).

Proof of (A.6). Consider

$$n(n-1)\ Es^2Y = E\ \Sigma \frac{1}{X_i}\ \Sigma\ (X_j-\bar{X})^2$$

$$= E\ \Sigma \frac{1}{X_i}\ (\Sigma X_j^2 - n\bar{X}^2)$$

$$= E\Sigma\Sigma \frac{X_j^2}{X_i} - \frac{1}{n}\ \Sigma\Sigma\Sigma \frac{X_iX_j}{X_k}$$

$$= nEX + n(n-1)\ EX^2E\frac{1}{X} - EX - 2(n-1)\ EX$$
$$- (n-1)\ EX^2E\frac{1}{X} - (n-1)(n-2)\ E^2XE\frac{1}{X}$$

and divide through by (n-1) to get

$$nEs^2Y = -EX + (n-1)\ EX^2E\ \frac{1}{X} - (n-2)\ E^2XE\ \frac{1}{X}$$

$$= -m_1 + (n-1)m_2E\ \frac{1}{X} - (n-2)m_1^2E\ \frac{1}{X}$$

Then subtract $n(m_2-m_1^2)E\ \frac{1}{X}$ to obtain the first version of (A.6). The second version follows from (A.8).

Proof of (A.7). Follows from (A.6) because s^2 and μ_2 are invariant under translations.

Proof of (A.1). Let us work with

$$E\bar{X}s^2 = \frac{1}{n(n-1)}\ E\{\Sigma X_i^2 \Sigma X_j - \frac{1}{n}\ (\Sigma X_i)^3\}$$

$$= \frac{1}{n(n-1)}\ \{nEX^3 + n(n-1)EX^2EX - EX^3 - 3(n-1)EX^2EX - (n-1)(n-2)E^3X\}$$

$$= \frac{1}{n}\ \{EX^3 + (n-3)EX^2EX - (n-2)E^3X\}$$

$$= \frac{1}{n}\ \{m_3 + (n-3)m_2m_1 - (n-2)m_1^3\}$$

and subtract $m_2m_1 - m_1^3$. The second version follows from (A.8).

Proof of (A.2). Let us begin with

$$(n-1)^2Es^4 = E(\Sigma X_i^2)^2 - \frac{2}{n}\ E\Sigma X_i^2(\Sigma X_j)^2 + \frac{1}{n^2}\ E(\Sigma X_i)^4\ ,$$

and treat the three main ingredients separately.

$$E(\Sigma X_i^2)^2 = nEX^4 + n(n-1)E^2X^2 \tag{A.9}$$

$$EX_i^2(\Sigma X_j)^2 = nEX^4 + n(n-1)E^2X^2 + n(n-1)(n-2)EX^2E^2X + 2n(n-1)EX^3EX \tag{A.10}$$

$$E(\Sigma X_i)^4 = nEX^4 + 4n(n-1)EX^3EX + 3n(n-1)E^2X^2 + 6n(n-1)(n-2)EX^2E^2X + n(n-1)(n-2)(n-3)E^4X \tag{A.11}$$

The proper combination of (A.9), (A.10), and (A.11) produces

$$Es^4 = \frac{1}{n}\{EX^4 - 4EX^3EX + \frac{1}{n-1}[(n^2-2+3)E^2X^2-2(n-2)(n-3)EX^2E^2X+ (n-2)(n-3)E^4x]\}$$

The subtraction of σ^4 in the form

$$\sigma^4 = E^2X^2 - 2EX^2E^2X + E^4X$$

yields

$$nVs^2 = EX^4 - 4EX^3EX + 3E^2X^2 - 4\sigma^4 + \frac{2}{n-1}\sigma^4$$

and this is (A.2). The second version follows from using (A.8) and modifying.

<u>Proof of (A.3)</u>. The variance of this form of the sample variance can be developed from (A.2) using (A.8). It also appears in [7, p. 183].

APPENDIX B

Moments of the Poisson

The Poisson random variable X had density

$$f(x;\lambda) = e^{-\lambda}\,\lambda^{x}/x!\ , \qquad x = 0,1,2,... \tag{B.1}$$

and it is well known that $\mu = \lambda$, $\sigma^2 = \lambda$. The probability generating function is

$$G(u) = E(u^X) = e^{-\lambda(1-u)} \tag{B.2}$$

and the moment generating function can be obtained from (B.2) by the replacement e^u for u. Moments can be obtained by repeated differentiation. We record

$$\begin{aligned} E(X) &= \lambda \\ E(X^2) &= \lambda^2 + \lambda \\ E(X^3) &= \lambda + 3\lambda^2 + \lambda^3 \\ E(X^4) &= \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^3 \end{aligned} \tag{B.3}$$

Use of (B.3) into (A.2) produces

$$\mu_3 = \lambda + 3\lambda^2 \tag{B.4}$$

$$\mathrm{Var}(s^2) = \frac{1}{n}\,\{2\lambda^2 + \lambda\} + o(\tfrac{1}{n}) \tag{B.5}$$

$$\mathrm{Cov}(\bar{X},s^2) = \frac{1}{n}\,\lambda \tag{B.6}$$

The moments of $(1+X)^{-1}$ can be obtained by integrating the generating function (B.2)

$$E(\frac{1}{1+X}) = E\int_0^1 u^X du = \int_0^1 G(u)\, du = e^{-\lambda}\int_0^1 e^{\lambda u}\, du = \frac{1}{\lambda}(1-e^{-\lambda}) \quad \text{(B.7)}$$

$$E(\frac{1}{1+X})^2 = E\int_0^1\int_0^1 u^X v^X\, du\, dv = \iint G(uv)\, du\, dv = e^{-\lambda}\iint e^{\lambda uv}\, du\, dv$$

$$= \frac{e^{-\lambda}}{\lambda}\int_0^1 \frac{e^{\lambda u}-1}{u}\, du = \frac{e^{-\lambda}}{\lambda}\sum_1^\infty \int_0^1 \frac{\lambda^j u^{j-1}}{j!}\, du$$

$$= e^{-\lambda}\sum_1^\infty \frac{\lambda^{j-1}}{j\cdot j!} \quad \text{(B.8)}$$

This opportunity is taken to record an alternative way of obtaining moments, (B.2), which in this case is somewhat easier than the differentiation of the moment generating function. One begins with the generating function $G(u) = E(u^X)$ and replaces the argument u by a product of dummy variable $uv \cdots z$ containing as many factors as the order of the moment to be calculated. Then one takes a partial derivative with respect to each variable u, v, etc. and the desired moment is obtained when each variable is set to unity. For example, $E(X^2)$ can be obtained from

$$G_{uv} = \frac{\partial^2}{\partial u\, \partial v} G(uv) = \frac{\partial^2}{\partial u\, \partial v} E(uv)^X = E(X^2 u^{X-1} v^{X-1}) \quad \text{(B.9)}$$

The four moments (B.3) can be obtained in this way replacing u with $uvwz$ in (B.2). The resulting derivatives are

$$G_u = \lambda G$$

$$G_{uv} = \lambda G[1 + \lambda uvwz]$$

$$G_{uvw} = \lambda G\{z[1 + \lambda uvwz] + \lambda uvwz^2[1 + \lambda uvwz] + \lambda uvwz^2\}$$

$$G_{uvwz} = \lambda uvw\, G_{uvw} + \lambda G\{[1 + \lambda uvwz] + z\lambda uvw + 2z\lambda uvw(1+\lambda uvwz) + \lambda^2 u^2 v^2 w^2 z^2 + 2z\lambda uvw\}$$

APPENDIX C

Moments of the Gamma

The gamma random variable X has density

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \tag{C.1}$$

and it is well-known that

$$\mu = \alpha\beta \qquad \sigma^2 = \alpha\beta^2 \tag{C.2}$$

Direct integration produces the formula, for $r > 0$

$$E(X^r) = \beta^r \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \tag{C.3}$$

and for $r < \alpha$

$$E(X^{-r}) = \frac{1}{\beta^2} \frac{\Gamma(\alpha-r)}{\Gamma(\alpha)} \tag{C.4}$$

Use of (C.3) in (A.1) and (A.2) produces

$$\mathrm{Var}(\bar{X}) = \frac{1}{n} \alpha\beta^2 \tag{C.5}$$

$$\mathrm{Cov}(\bar{X},s^2) = \frac{2}{n} \alpha\beta^3 \tag{C.6}$$

$$\mathrm{Var}(s^2) = 2\alpha\beta^4 \left(\frac{\alpha}{n-1} + \frac{3}{n}\right) \tag{C.7}$$

Using (C.4) one readily calculates

$$E\left(\frac{1}{X}\right) = \frac{1}{\beta(\alpha-1)} \qquad 1 < \alpha \tag{C.8}$$

$$\mathrm{Var}\left(\frac{1}{X}\right) = \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)} \qquad 2 < \alpha \tag{C.9}$$

and letting $Y = \frac{1}{n}\sum_1^n (1/X_i)$, we find that, with the aid of (A.4),

$$\mathrm{Cov}(\bar{X},Y) = \frac{-1}{n(\alpha-1)} \qquad 1 < \alpha \tag{C.10}$$

and with the aid of (A.6)

$$\mathrm{Cov}(s^2,Y) = 0 \tag{C.11}$$

Because of the formula

$$\Gamma'(y) = \int_0^\infty \ln(y)\ y^{\alpha-1}\ e^{-y}\ dy \tag{C.12}$$

one can develop, for $r > -\alpha$,

$$E(X^r \ln X) = \beta^r \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} (\ln \beta + \psi(\alpha+r)) \tag{C.13}$$

since $\Gamma'(\alpha) = \Gamma(\alpha)\ \psi(\alpha)$.

APPENDIX D

Moments of the Negative Binomial

The density has the form

$$f(x;,r,p) = \frac{\Gamma(r+x)}{x!\Gamma(r)} q^x p^r \qquad \text{for } x = 0,1,\ldots \tag{D.1}$$
$$0 < r,\ 0 < p < 1,\ p+q = 1$$

Its probability generating function

$$G(u) = E(u^X) = \frac{p^r}{(1-qu)^r} \tag{D.2}$$

will be exploited broadly. Successive derivatives of (D.2) evaluated at $u = 1$ produce the factorial moments

$$m^{(s)} = EX(X-1) \cdots (X-s+1) = (q/p)^s\, r(r+1) \cdots (r+s-1) \tag{D.3}$$

to which one may apply some orderly substitution and obtain the first four moments, using $A = q/p$

$$\begin{aligned} EX &= Ar \\ EX^2 &= A^2 r(r+1) + Ar \\ EX^3 &= A^3 r(r+1)(r+2) + 3A^2 r(r+1) + Ar \\ EX^4 &= A^4 r(r+1)(r+2)(r+3) + 5A^3 r(r+1)(r+2) + 4A^2 r(r+1) + Ar \end{aligned} \tag{D.4}$$

The mean and variance of the population are

$$\mu = rq/p \qquad\qquad \sigma^2 = rq/p^2 \tag{D.5}$$

Use of (D.4) in (A.1) and (A.2) provides the covariance matrix for the ordinary method of moments

$$\begin{aligned}
n\,\mathrm{Var}(\bar{X}) &= A^2r + Ar = rq/p^2 = \sigma^2 \\
n\,\mathrm{Cov}(\bar{X},s^2) &= 2A^3r + 3A^2r + Ar = \sigma^2\,\frac{1+q}{p} \\
n\,\mathrm{Var}(s^2) &= 2A^4r(r+3) + 4A^3r(r+3) + A^2r(2r+7) + Ar \\
&= \sigma^2[2(r+3)q+p^2]/p^2
\end{aligned} \tag{D.6}$$

The harmonic mean alternative requires

$$E\left(\frac{1}{1+X}\right) = \frac{p-p^r}{(r-1)q} \tag{D.7}$$

$$E\left(\frac{1}{1+X}\right)^2 = \frac{p^r}{(r-1)q}\int_0^1 \frac{1}{u}\left[\frac{1}{(1-qu)^{r-1}} - 1\right] du \tag{D.8}$$

for $r \neq 1$. The case $r = 1$ is treated in [12, Appendix A]. Expressions (D.7) and (D.8) are justified next along with computational formulae for (D.8) when r is either a whole number or a whole number plus 0.5.

<u>Proof of (D.7)</u>. The generating function (D.2) is integrated.

$$E\left(\frac{1}{1+X}\right) = \int_0^1 G(u)\,du = p^r\int_0^1 \frac{du}{(1-qu)^2} = \left.\frac{-p^r}{(r-1)(-q)(1-qu)^{r-1}}\right|_{n=0}^{n=1} = \frac{p-p^r}{(r-1)q}$$

as required.

<u>Proof of (D.8)</u>. Replace u by uv and integrate twice.

$$E\left(\frac{1}{1+X}\right)^2 = \int_0^1\int_0^1 G(uv)\,du\,dv = \int_0^1 dv \left\{ \frac{p^r}{(r-1)qu} \frac{1}{(1-quv)^{r-1}} \Bigg|_{u=0}^{u=1} \right\}$$

$$= \frac{p^r}{(r-1)q} \int_0^1 \frac{1}{u}\left[\frac{1}{(1-qu)^{r-1}} - 1\right] du$$

as required.

Computational formulae require managing integrals of the type

$$J_r = \int_0^1 \left[\frac{1}{u(1-qu)^r} - \frac{1}{u}\right] du \tag{D.9}$$

Clearly $J_0 = 0$. It is useful to apply the partial fraction representation repeatedly.

$$\frac{1}{u(1-qu)^r} = \frac{q}{(1-qu)^r} + \frac{1}{u(1-qu)^{r-1}} \tag{D.10}$$

Since

$$\int_0^1 \frac{q\,du}{(1-qu)^r} = \frac{1}{r-1}\left[\frac{1}{p^{r-1}} - 1\right]$$

we have, from (D.10),

$$J_r = J_{r-1} + \frac{1}{r-1}\left[\frac{1}{p^{r-1}} - 1\right]$$

Let $\langle r\rangle$ be the greatest integer in r and apply the above formula to get the representation

$$J_r = \sum_{j=1}^{\langle r\rangle} \frac{1}{r-j}\left[\frac{1}{p^{r-j}} - 1\right] + J_{r-\langle r\rangle} \tag{D.11}$$

which is valid for $r > \langle r\rangle$.

Let $r = \langle r \rangle$. One can show directly from (D.9) that

$$J_1 = \int_0^1 \frac{q\,du}{(1-qu)} = -\ln p \tag{D.12}$$

and then

$$J_r = \sum_{j=1}^{r-1} \frac{1}{r-j}\left[\frac{1}{p^{r-j}} - 1\right] - \ln p = \frac{1}{p^r}\sum_{j=1}^{r-1} \frac{1}{j}\,[p^{r-j}-p^r] - \ln p \tag{D.13}$$

Accordingly it is recognized from (D.8) and (D.9) that $E(1/1+X)^2 = [p^r/(r-1)q]J_{r-1}$ which, together with (D.7) and (D.13) enables

$$\mathrm{Var}(\frac{1}{1+X}) = \frac{1}{(r-1)q}\left\{\sum_{j=1}^{r-2} \frac{1}{j}\,[p^{r-j}-p^r] - p^r \ln p - \frac{(p-p^r)^2}{(r-1)q}\right\} \tag{D.14}$$

for r an integer ≥ 2 (empty sum is zero), and for $r = 1$,

$$\mathrm{Var}(\frac{1}{1+X}) = \frac{p}{q}\left\{\sum_{j-1}^{\infty} \frac{q^j}{j^2} - \frac{p}{q}(\ln p)^2\right\} \tag{D.15}$$

This latter formula is developed in [12], see (A.3) and (A.5).

Let $r = \langle r \rangle + .5$. The exploitation of (D.11) requires dealing with

$$J_{r-\langle r \rangle} = J_{1/2} = \int_0^1 \left[\frac{1}{u\sqrt{1-qu}} - \frac{1}{u}\right] du\ .$$

Making the change $qu = \sin^2\theta$ provides for manageable integrals of trigonometric functions. See [14, p. 316]. The result is

$$J_{1/2} = 2\ln\frac{2}{1+\sqrt{p}} \tag{D.16}$$

and

$$E\left(\frac{1}{1+X}\right)^2 = \frac{1}{(r-1)q}\left\{\sum_{j=2}^{<r>} \frac{1}{r-j}\left[p^j - p^r\right] + p^r J_{1/2}\right\} \tag{D.17}$$

which is valid for $<r> \geq 1$. For $r = 1/2$ one needs, from (D.9),

$$J_{-1/2} = \int_0^1 \left[\frac{\sqrt{1-qu}}{u} - \frac{1}{u}\right] du$$

$$= 2\sqrt{1-qu}\,\Big|_{u=0}^{u=1} + \int_0^1 \left[\frac{1}{u\sqrt{1-qu}} - \frac{1}{u}\right] du$$

$$= 2\sqrt{p} - 2 + J_{1/2} \tag{D.18}$$

using formula [14, p.316]. From (D.8)

$$E\left(\frac{1}{1+X}\right)^2 = -\frac{2}{q}\sqrt{p}\,\{2\sqrt{p} - 2 + J_{1/2}\} = \frac{4}{q}\left\{\sqrt{p} - p - \sqrt{p}\,\ln\frac{2}{1+\sqrt{p}}\right\} \tag{D.19}$$

for $r = 1/2$.

The information matrix (4.19) contains a difficult element Λ_{22}, (4.20). It may be managed using an integral representation of the trigamma function [1, p. 259],

$$\psi'(r) = -\int_0^1 \frac{\ln u}{1-u} u^{r-1}\, du \tag{D.20}$$

It follows that

$$\Lambda_{22} = \psi'(r) - E\{\psi'(r+X)\}$$

$$= -\int_0^1 \frac{\ln u}{1-u}\left[u^{r-1} - E(u^{r-1+X})\right] du$$

$$= -\int_0^1 \frac{\ln u}{1-u} u^{r-1}\left[1 - \frac{p^r}{(1-qu)^r}\, r\right] du \tag{D.21}$$

This relationship can be expressed better if we make the change of variable

$$w = pu/(1-qu)$$

and manipulate to obtain

$$\Lambda_{22} = - \int_0^1 \frac{\ln(p+qw)}{1 - w} w^{r-1} dw \tag{D.22}$$

There is advantage in using (D.22) in the development of the determinant of Λ, (4.19),

$$|\Lambda| = - \frac{r}{qp^2} \int_0^1 \frac{\ln(p+qw)}{1 - w} w^{r-1} dw - \frac{1}{p^2}$$

$$= \frac{1}{p^2} \{- \frac{1}{q} \int_0^1 \frac{\ln(p+qw)}{1 - w} d(w^r) - 1\}$$

$$= \frac{1}{qp^2} \int_0^1 w^r d(\frac{\ln(p+q+w)}{1 - w}) \tag{D.23}$$

using partial integration and

$$\frac{\ln(p+qw)}{1 - w} = - \sum_1^{\infty} \frac{q^n}{n} (1-w)^{n-1}$$

In this form it is easily checked that

$$\lim_{r \to 0} |\Lambda| = \frac{1}{qp^2} \int_0^1 d(\frac{\ln(p+qw)}{1 - w}) dw = \frac{-\ln(p)-q}{qp^2}$$

$$\lim_{r \to \infty} |\Lambda| = 0 \tag{D.24}$$

For computational purposes one can exploit the expansion

$$\frac{\ln(p+qw)}{1-w}) = \sum_{n=2}^{\infty} \frac{q^n}{n} (n-1)(1-w)^{n-2} \, dw$$

which, when used in (D.23), produces

$$|\Lambda| = \frac{1}{qp^2} \sum_{2}^{\infty} q^n \frac{n-1}{n} \int_0^1 w^r (1-w)^{n-2} \, dw$$

$$= \frac{1}{p^2} \sum_{1}^{\infty} q^n \frac{n}{n+1} \frac{r!(n-1)!}{(n+r)!} \tag{D.25}$$

$$= \frac{1}{p^2} \sum_{n=1}^{\infty} \frac{q^n}{n+1} \frac{r!\, n!}{(n+r)!}$$

Similar efforts applied to (D.22) can produce

$$\Lambda_{22} = \sum_{n=1}^{\infty} \frac{q^n}{n^2} \frac{(r-1)!\, n!}{(r+n-1)!} \tag{D.26}$$

APPENDIX E

Moments of the Beta Distribution

The Beta random variable X has density

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

$$\text{for} \quad 0 \le x \le 1,\ 0 < \alpha,\ 0 < \beta \tag{E.1}$$

and

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx \ .$$

This is called the $B(\alpha,\beta)$ distribution and it is useful to note that $1-X$ has a $B(\beta,\alpha)$ distribution.

Moments are obtained directly by manipulating gamma and beta functions. For $r > 0$

$$E(X^r) = \frac{\Gamma(\alpha+\beta)\ \Gamma(\alpha+r)}{\Gamma(\alpha)\ \Gamma(\alpha+\beta+r)} = \frac{(\alpha+\beta-1)!\ (\alpha+r-1)!}{(\alpha-1)!\ (\alpha+\beta+r-1)!} \tag{E.2}$$

$$E[X^r(1-X)^s] = \frac{\Gamma(\alpha+\beta)\ \Gamma(\alpha+r)\ \Gamma(\beta+s)}{\Gamma(\alpha)\ \Gamma(\beta)\ \Gamma(\alpha+\beta+r+s)} \tag{E.3}$$

and the mean and variance follow

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{E.4}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\ (\alpha+\beta+1)} \tag{E.5}$$

Moments of negative order exist if α is large enough. If $\alpha > r,\ \beta > s$ then

$$E(X^{-r}) = \frac{\Gamma(\alpha+\beta)\ \Gamma(\alpha-r)}{\Gamma(\alpha)\ \Gamma(\alpha+\beta-r)} \qquad (E.5)$$

$$E(X^{-r})(1-X)^{-s} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)} \frac{\Gamma(\alpha-r)\ \Gamma(\beta-s)}{\Gamma(\alpha+\beta-r-s)} \qquad (E.7)$$

and the harmonic mean option will be available,

$$E(\frac{1}{X}) = 1 + \frac{\beta}{\alpha-1} \qquad 1 < \alpha \qquad (E.8)$$

$$Var(\frac{1}{X}) = \frac{\beta(\alpha+\beta-1)}{(\alpha-1)^2\ (\alpha-2)} \qquad 2 < \alpha \qquad (E.9)$$

and if $\alpha > 1,\ \beta > 1$

$$Cov(\frac{1}{X}, \frac{1}{1-X}) = -\frac{(\alpha+\beta-1)}{(\alpha-1)\ (\beta-1)} \qquad (E.10)$$

Symmetric Beta

Set $\alpha = \beta$ and obtain

$$\mu = 1/2 \qquad (E.11)$$

$$\sigma^2 = \frac{1}{4(2\alpha+1)} \qquad (E.12)$$

$$E(\frac{1}{X}) = \frac{2\alpha-1}{\alpha-1} \qquad 1 < \alpha \qquad (E.13)$$

$$Var(\frac{1}{X}) = \frac{\alpha(2\alpha-1)}{(\alpha-1)^2\ (\alpha-2)} \qquad 2 < \alpha \qquad (E.14)$$

$$Cov(\frac{1}{X}, \frac{1}{1-X}) = -\frac{(2\alpha-1)}{(\alpha-1)^2} \qquad 1 < \alpha \qquad (E.15)$$

APPENDIX F

APL PROGRAMS

The function HARP performs the iteration (3.7) to estimate the Poisson parameter from the harmonic mean. The left argument X is the initial value (usually $\bar{x}$) and the right argument is y of (3.6). The function HMV computes the variance of y using (B.7) and (B.8). The right argument is the set of parameter values, λ.

```
    ∇ L←X HARP Y;LL
[1]   L←X
[2]  L1:LL←L
[3]   L←(1-*-L)÷Y
[4]   L
[5]   →L1×ι(|L-LL)≥0.0001
    ∇
```

```
    ∇ V←HMV L;B;D;M;N
[1]   N←50
[2]   V←(L÷*1)∘.*ιN
[3]   D←((M←1↑ρV),N)ρ(*-ιN)×(ιN)×!ιN
[4]   B←⍉(N,M)ρ*-L
[5]   V←+/B×V÷D
[6]   V←V÷L
[7]   V←V-((1-*-L)÷L)*2
    ∇
```

The function MONT produces $E(X^r)$ for the symmetric beta distribution using (E.2) with $\alpha = \beta$. The left argument is the (integral) order of the moment and the right argument is the parameter. The function VAR computes the variance of the estimator, $\tilde{\alpha}$ of (3.16) using (3.17) and (A.2). Again the argument is the parameter

```
    ∇ Z←R MONT A;N;I
      I←0
      Z←(N,N←ρA)ρ1
    L1:I←I+1
      Z←Z×(⍉(N,N)ρA+I-1)÷A∘.+A+I-1
      →L1×⍳I<R
    ∇
```

```
    ∇ V←VAR A
      V←(4 MONT A)-(4×(3 MONT A)×(1 MONT A))-(3×(2 MONT A)*2)-4×(4+8×A
      V←4×V×(1+2×A)*4
    ∇
```

The polygamma functions are computed using PSI and JEX. When the (scalar) left argument, N, is zero the psi function is produced. Integral values of N index the order of the derivative of psi. The argument of the function appears on the right. The technique comes from the asymptotic expansions in Abramowitz.

```
    ∇ P←N PSI Y;C;IV;JIV;K;KK;YY;V;Z;T;I
      ⍝ N IS THE ORDER OF THE DERIVATIVE OF THE PSI FUNCTION
      ⍝ Y (>0) IS THE ARGUMENT,SCALAR OR VECTOR
      C←10
      IV←⍳ρY←,Y
      P←Z←K←(ρY)ρ0
      KK←⌈C-Y[JIV←(V←Y<C)/IV]
      →L1×⍳(ρIV)=ρ(~V)/IV
      T←Y[JIV]
      I←0
    L2:I←I+1
      YY←KK[I]ρT[I]
      Z[I]←(!N)×+/((YY-1)+⍳KK[I])*-1+N
      →L2×⍳I<ρJIV
      Z←V\Z[⍳ρJIV]
      K←V\KK
    L1:→S1×⍳N>0
      P←-(⍟K+Y)-(2×K+Y)*-1
      →2+⌶26
    S1:P←((!N-1)×(Y+K)*-N)+(!N)×0.5×(Y+K)*-N+1
      P←((¯1)*N+1)×P+Z+N JEX Y+K
    ∇
```

```
    ∇ J←N JEX Y;C;M;F;E;A
[1]   ⍝ USED IN THE POLYGAMMA FUNCTIONS
[2]   C←-C←(5,(7÷5),(5÷7),(11÷25),(5×455÷11×691),(691÷7×455))
[3]   C←((ρY←,Y),6)ρC
[4]   M←7
[5]   F←(Y*2)∘.×(2+2×⍳M-1)×(1+2×⍳M-1)÷(N+2×⍳M-1)×(N+1+2×⍳M-1)
[6]   E←C×F
[7]   A←(7÷6)×(Y*-N+2×M)×(!¯1+N+2×M)÷(!2×M)
[8]   J←A×1+E[;6]×1+E[;5]×1+E[;4]×1+E[;3]×1+E[;2]×1+E[;1]
    ∇
```

The efficiency of the usual moment estimator (4.5) of gamma distribution parameters is produced by EFF using (4.8) and (4.14). The efficiency of (4.10) is computed by EFFH using (4.13) and (4.14), and the relative efficiency (4.15) is the function REFF. Only the parameter α is needed and it is entered on the right for all three. It must be > 2 for the latter two functions.

```
    ∇ E←EFF Y
[1]   E←(2×(Y+1)×((Y×(1 PSI Y))-1))*-1
    ∇
```

```
    ∇ E←EFFH Y
[1]   E←2÷(Y-2)÷(Y-1)*2
[2]   E←E×(Y×1 PSI Y)-1
[3]   E←E*-1
    ∇
```

```
    ∇ Z←REFF A
[1]   Z←(A-1)×(A-1)÷(A+1)×A-2
    ∇
```

We turn now to the programs that support the negative binomial distribution. The function MM computes the ordinary method of moments estimators $\tilde{p}$, $\tilde{r}$ of (4.25). The left argument is $\bar{x}$ and the right argument is s^2.

```
      ∇ M←X MM S
[1]     P←X÷S
[2]     R←(X*2)÷S-X
      ∇
```

The determinant of the asymptotic covariance matrix of $\tilde{p}$, $\tilde{r}$ is produced (see (4.29)) by DM, whose left and right arguments are r and p.

The function PSIB computes $\bar{x}$ (= XB), s^2 (= S), y (= Y), the estimators $\tilde{p}$, $\tilde{r}$ (using MM), and the maximum likelihood estimate $\hat{p}$, $\hat{r}$ by applying Newton's method to (4.18) using $\tilde{p}$, $\tilde{r}$ for initialization. The left argument, F, is the vector of observed frequencies corresponding to the right argument J, the vector of variate values.

```
      ∇ Z←F PSIB J;PH;PHD;Z;ZZ
[1]     XB←(+/J×F)÷+/F
[2]     S←(÷(+/F)-1)×(+/F×J*2)-(+/F)×XB*2
[3]     Y←(+/F÷J+1)÷+/F
[4]     XB MM S
[5]     R,P
[6]    L1:PP←P
[7]     Z←+/(0 PSI R+J)×F÷+/F
[8]     ZZ←+/(1 PSI R+J)×F÷+/F
[9]     PH←Z-(0 PSI R)-(⍟R)-⍟R+XB
[10]    PHD←ZZ-(1 PSI R)-(÷R)-÷R+XB
[11]    R←R-PH÷PHD
[12]    P←R÷R+XB
[13]    R,P
[14]    →L1×⍳(|PP-P)≥0.0001
      ∇
```

The function INF1 computes the difficult element of the information matrix, Λ_{22} of (4.20), using (D.26). The left and right arguments are r and p, resp. The vector r must be whole numbers except it will also handle the values .5, 1.5, ... , 4.5.

```
     ∇ Z←R INF1 P;N;A;B;C;ZZ;J;V;N1;NN
[1]    N←⍳M
[2]    C←(1-P←,P)∘.*N
[3]    A←((⍴N),⍴R←,R)⍴0
[4]    J←0
[5]   L1:J←J+1
[6]    V←(N+R[J])>55
[7]    →L2×⍳(⍴N)=⍴NN←(~V)/N
[8]    ZZ←(⍴N1←V/N)⍴0
[9]    A[N1;J]←×/(ZZ∘.+⍳R[J]-1)÷N1∘.+⍳R[J]-1
[10]   →L3×⍳(⍴NN)=0
[11]  L2:A[NN;J]←(!NN)×(!R[J]-1)÷!NN+R[J]-1
[12]  L3:→L1×⍳J<⍴R
[13]   B←⍉((⍴R),⍴N)⍴N*2
[14]   Z←C+.×A÷B
     ∇
```

The function DETI and DI both compute the determinant of the information matrix, but the former uses INF1 in (4.19) and can handle the same set of r values which are whole numbers plus .5. The latter, DI, uses (D.25) and all values of r must be whole numbers.

```
     ∇ V←R DETI P
[1]    V←(R INF1 P)×⍉(R←,R)∘.÷(P*2)×1-P
[2]    V←V-⍉((⍴R),⍴P)⍴P*-2
     ∇
```

```
∇ D←R DI P;A;B;C;N;P1;J;Z
  N←⍳M̲
  C←(1-P←,P)∘.*N
  P1←⍉((ρR←,R),ρP)ρP*-2
  A←((ρN),ρR←,R)ρ0
  Z←(ρN)ρ0
  J←0
L1:J←J+1
  A[;J]←×/(Z∘.+⍳R[J])÷N∘.+⍳R[J]
  →L1×⍳J<ρR
  B←⍉((ρR),ρN)ρN+1
  D←P1×D←C+.×A÷B
∇
```

The determinant (4.29) of the asymptotic covariance matrix (4.28) is computed by the function DM

```
∇ D←R DM P
  D←2×(1+R←,R)∘.×(P*2)÷1-P←,P
∇
```

and the efficiency of the moment estimator is NBEF

```
∇ E←R NBEF P
  E←÷(R DM P)×⍉R DETI P
∇
```

The function HAR implements the iteration scheme described in (4.32) and produces the harmonic mean based alternative estimators p^*, r^* from the system (4.30). The left argument is $\bar{x}$ and the right argument is y.

```
∇ P←X HAR Y;PP;C;G;GP;Z
  P←Q←0.5
  C←Y-1-Y×X
  R←X
L1:PP←P
  G←(Z←P*R)-Y-C×P
  GP←C+Z×X×(Q+⍟P)÷Q*2
  P←P-G÷GP
  R←X×P÷Q←1-P
  R,P,G
  →L1×⍳(|G)≥1E¯6
∇
```

The Var(1/1+X) is computed by VHM1 using (D.14) and (D.15) for r values that are whole numbers, and using (D.17) and (D.19) for r values that are half way between whole numbers.

```
    ∇ Z←R VHM1 P;AA;BB;B;I;PP;RR;S;H;HH;HHH;M;U1;U2;J;N;NN;Q;SS
[1]    ⍝ VAR OF ÷(1+X) FOR NEG BIN(R,P);R MUST BE WHOLE OR WHOLE PLUS .5
[2]    R←(~U2←R=0.5)/R←(~U1←R=1)/R
[3]    Z←S←((RR←ρR←,R),PP←ρP←,P)ρ0
[4]    →L3×ι(ρR)=0
[5]    I←0
[6]   L1:I←I+1
[7]    B←R[I]-1
[8]    →(3+⌶26)×ιB≠⌊B
[9]    B←B-1
[10]   S[I;]←-⍟P
[11]   →(2+⌶26)×ιB=⌊B
[12]   S[I;]←2×⍟2÷1+P*0.5
[13]   BB←(PP,⌊B)ρR[I]-1+ι⌊B
[14]   AA←(P∘.*-R[I]-1+ι⌊B)-1
[15]   S[I;]←S[I;]++/AA÷BB
[16]   →L1×ιI<⌊RR←ρR
[17] L3:R←(~U2)\R
[18]   →L4×ι(ρR)=0
[19]   R[U2/ιRR←RR++/U2]←0.5
[20]   S←(~U2)\[1] S
[21]   →(2+⌶26)×ι0=+/U2
[22]   S[U2/ιRR;]←+(2×⍟2÷1+P*0.5)+2×(P*0.5)-1
[23]   Z←S×(H←⍉P∘.*R)÷HHH←(R-1)∘.×1-P
[24]   M←((HH←(RR,PP)ρP)-H)÷HHH
[25]   Z←Z-M*2
[26] L4:N←(2×⍟50÷P×1000)÷⍟Q←1-P
[27]   N[(V←N≤50)/ιPP]←50
[28]   SS←PPρ0
[29]   J←0
[30] L2:J←J+1
[31]   SS←+/(Q∘.*ιNN)÷(PP,NN)ρ(ιNN←⌈N[J])*2
[32]   →L2×ιJ<PP
[33]   →(1+⌶26)×(+/U1)≠0
[34]   R←(~U1)\R
[35]   R[U1/ιRR←RR++/U1]←1
[36]   Z←(~U1)\[1] Z
[37]   Z[U1/ιRR;]←P×(SS-P×((⍟P)*2)÷Q)÷Q
    ∇
```

The function DMSI computes the determinant of M^{*-1} from (4.36). Two auxiliary functions are needed: G22 provides g_{21} and g_{22} from the second row of A^{*} in (4.33) (compare (2.17)) and DMS1 is needed to handle values of $r = 1$, special handling being required becuase of (4.34).

```
    ∇ X←R DMSI P;RR;PP;HH;C;L;G;G21;Z;U
[1]   X←((ρR),ρP)ρ0
[2]   R←(~U←R=1)/R
[3]   Z←(L←R∘.÷(P*2)×Q←1-P)×G←⍉R G22 P
[4]   Z←(Z+(⍉G21)÷HH←((RR←ρR),PP←ρP)ρP)*2
[5]   Z←Z÷(L×C←(((R-1)∘.×Q)*2)×R VHM1 P)-⍉G21*2
[6]   X[(~U)/⍳RR++/U;]←Z
[7]   →(2+⌶26)×⍳0=+/U
[8]   X[U/⍳RR++/U;]←DMS1 P
    ∇
```

```
    ∇ Z←DMS1 P;W
[1]   Z←((0.5×(⍟P)*2)+W←1-P-⍟P)*2
[2]   Z←Z÷(((ρP)ρ1 VHM1 P)×((1-P)*3))-(W×P)*2
    ∇
```

```
    ∇ Z←R G22 P;PP;RR;T;TT;H;HH;V;W
[1]   H←((PP←ρP←,P),RR←ρR←,R)ρP∘.*R
[2]   HH←⍉(RR,PP)ρP
[3]   Z←HH-H
[4]   V←R=1
[5]   TT←(~V)/⍳RR
[6]   T←H×⍉(RR,PP)ρ⍟P
[7]   Z[;TT]←T[;TT]+Z[;TT]÷(ρZ[;TT])ρ(~V)/R-1
[8]   G21←((PP,RR)ρR)×H÷HH
[9]   G21←G21-(1-H)÷⍉(RR,PP)ρ1-P
    ∇
```

The efficiency of the harmonic mean based estimator p^*, r^* is computed by the function EFF

```
      ∇ E←R EFF P
[1]     E←(R DMSI P)⌹R DETI P
      ∇
```

Both EFF and NBEF used DETI and, for that reason, are restricted to only a few fractional values of r. The relative efficiency of p^*, r^* with respect to $\tilde{p}$, $\tilde{r}$ is not so restricted. It is computed by RELEF and accepts any $r > 0$.

```
      ∇ E←R RELEF P
[1]     E←((R DM P)×(R DMSI P))
      ∇
```

Methods for computing the efficiency of beta distribution estimators are supported by the following programs: The determinant of the information matrix (4.39) is computed by the functions DINF and DDINF. The former takes a single (vector) argument and produces a symmetric, square matrix of values $|\Lambda|$ for all pairs of components of the arguments. If the arguments α and β must be entered separately, then the two argument DDINF can be used.

```
      ∇ L←DINF A;N;B;C
[1]     C←,L←A∘.+A
[2]     B←(N,N←ρA)ρ1 PSI L
[3]     L←(C∘.×C)-(C∘.+C←1 PSI A)×B
      ∇
```

```
      ∇ L←A DDINF B;M;N;CA;CB;D
[1]     L←,L←A∘.+B
[2]     D←((M←ρA),N←ρB)ρ1 PSI L
[3]     L←(CA∘.×CB)-((CA←1 PSI A)∘.+CB←1 PSI B)×D
      ∇
```

The function DETMM accepts two by two matrices A (on the left) and C (on the right) and computed the determinant $|M| = |C|/|A|^2$ of (2.7). The function ADET uses it to produce an array of such determinants. The arguments H and C are four-dimensional and may be thought of as an M by N array of 2 by 2 matrices. The left set, H, are the coefficient arrays A of (2.2) and the right set, C, are the covariances (2.6)

```
     ∇ M←A DETMM C
[1]    M←(⌹A)+.×⍉C⌹A
[2]    M←(M[1;1]×M[2;2])-M[1;2]*2
     ∇
```

```
     ∇ MM←H ADET C;N;I;J;M
[1]    MM←((M←1↑¯2↑⍴H),N←¯1↑⍴H)⍴0
[2]    I←J←0
[3]   L1:I←I+1
[4]    J←0
[5]   L2:J←J+1
[6]    MM[I;J]←H[;;I;J] DETMM C[;;I;J]
[7]    →L2×⍳J<N
[8]    →L1×⍳I<M
     ∇
```

The computation of the efficiency of the ordinary moment estimator (4.42) requires the coefficient array (4.43) and the convariance array (4.45). The former is computed by the function COEFM and the latter by COVM. Each takes a single vector argument and computes the required values for all pairs of components in the argument. The function COVM requires the beta distribution moments (E.2) and these are computed by MONT.

```
    ∇ H←COEFM A;N;B
[1]   H←(2,2,N,N←ρA)ρ1
[2]   H[1;1;;]←(N,N)ρ-A
[3]   H[1;2;;]←⍉(N,N)ρA
[4]   B←(A∘.×A)×(A∘.+A)÷(A∘.+A+1)
[5]   H[2;1;;]←B+(A∘.-A)×(N,N)ρA
[6]   H[2;2;;]←B+((-A)∘.+A)×⍉(N,N)ρA
[7]   H←H÷(2,2,N,N)ρA∘.+A
    ∇
```

```
    ∇ Z←R MONT A;N;I
[1]   I←0
[2]   Z←(N,N←ρA)ρ1
[3]  L1:I←I+1
[4]   Z←Z×(⍉(N,N)ρA+I-1)÷A∘.+A+I-1
[5]   →L1×ιI<R
    ∇
```

All these are utilized by EFBM which computes the efficiency (2.22). The output is a symmetric matrix. The argument must have positive components.

```
    ∇ E←EFBM A;L;C;H;M
[1]   L←DINF A
[2]   C←COVM A.
[3]   H←COEFM A
[4]   M←H ADET C
[5]   E←÷M×L
    ∇
```

The efficiency of (4.48) is handled in similar fashion. (Also with single arguments.) An array of 2 by 2 matrices (4.49) is produced by the function COEF and the matching matrices (4.50) by COV. These are used by EFBH to compute a symmetric matrix of efficiencies. The argument must have all components > 2.

```
    ∇ C←COVM A;N;SS;B;S
[1]   C←(2,2,N,N←ρA)ρ1
[2]   C[1;1;;]←(A∘.×A)÷A∘.+A+1
[3]   S←C[1;1;;]÷(A∘.+A)*2
[4]   SS←((A∘.+A)*3)×A∘.+A+1
[5]   C[1;2;;]←C[2;1;;]←SS×B←(3 MONT A)-(3×(2 MONT A)×(1 MONT A))-2×(1 MONT A)*3
[6]   C[2;2;;]←(A∘.+A)×SS×(A∘.+A+1)×B←(4 MONT A)-(4×(3 MONT A)×(1 MONT A))-(3×(2 MONT A)*2)-4×S*2
    ∇
```

```
     ∇ H←COEF A;N
[1]    H←(2,2,N,N←⍴A)⍴-1
[2]    H[2;2;;]←A∘.÷A-1
[3]    H[1;1;;]←⍉H[2;2;;]
     ∇
```

```
     ∇ C←COV A;N
[1]    C←(2,2,N,N←⍴A)⍴1
[2]    C[1;1;;]←(A∘.+A-1)×⍉A∘.÷A-2
[3]    C[1;2;;]←C[2;1;;]←-A∘.+A-1
[4]    C[2;2;;]←(A∘.+A-1)×(A∘.÷A-2)
     ∇
```

```
     ∇ E←EFBH A;C;H;L
[1]    L←DINF A
[2]    H←COEF A
[3]    C←COV A
[4]    M←H ADET C
[5]    E←÷M×L
     ∇
```

The estimator (4.52) is managed in like fashion, only this time the arguments α (left) and β (right) must be entered separately with $\alpha > 2$, $\beta > 0$. The coefficients (4.53) are computed by the function COEFH and the covariances (4.54) by COVH. These are drawn on by EFBMH to produce the efficienices

```
     ∇ H←A COEFH B;M;N
[1]    H←(2,2,(M←⍴A),N←⍴B)⍴-1
[2]    H[1;1;;]←÷(A∘.+B)÷(M,N)⍴-B
[3]    H[1;2;;]←(⍉(N,M)⍴A)÷A∘.+B
[4]    H[2;1;;]←÷(A-1)∘.÷B
     ∇
```

```
∇ C←A COVH B;N;M
  C←(2,2,(M←ρA),N←ρB)ρ0
  C[1;1;;]←(A∘.×B)÷A∘.+B+1
  C[1;2;;]←C[2;1;;]←(M,N)ρ-B
  C[2;2;;]←(A∘.+B-1)÷(A-2)∘.÷B
∇
```

```
∇ E←A EFBMH B;C;H;M;L
  L←A DDINF B
  H←A COEFH B
  C←A COVH B
  M←H ADET C
  E←÷M×L
∇
```

REFERENCES

[1] Abramowitz, M. and Stegun, I.A., Handbook of Mathematical Functions, National Bureau of Standards, Applied Mathematics Series, 55 (1969).

[2] Bhattacharyya, A., "On Some Analoges of the Amount of Information and Their Use in Statistical Estimation," Sankya 8 (1946), p. 1

[3] Choi, S.C. and Wette, R., "Maximum Likelihood Estimation of the Parameters of a Gamma Distribution and Their Bias," Technometrics 11 (1969).

[4] Cramér, Harald, Mathematical Methods of Statistics, Princeton (1951), pp. 489-494.

[5] Feller, W., Introduction to Probability Theory with Applications, Vol. I, John Wiley and Sons, New York (1957).

[6] Graybill, F. A., Introduction to Matrices with Applications in Statistics, Wadsworth (1969).

[7] Hoel, P., Port, S. and Stone, C., Introduction to Statistical Theory, Houghton Mifflin (1971).

[8] Kendall, M.G., The Advanced Theory of Statistics, Vol. II, Hafner, 1959

[9] LeCam, L.M., "On the Asymptotic Theory of Estimation and Testing Hypotheses," Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, University of California Press, 1956.

[10] Lehman, E.L., "Notes on the Theory of Estimation," University of California, Department of Statistics, 1950.

[11] Rao, C.R., "Information and the Accuracy Attainable in the Estimation of Statistical Parameters," Bull. Calcutta Math. Soc. 37, No. 3 (1945), p. 81.

[12] Read, R.R., "Methodology for Efficiency and Alteration of the Likelihood System," USNPS Technical Report NPS55-77-9, Monterey, Calif. (1977).

[13] Reep, C., Pollard, R., Benjamin, B., "Skill and Chance in Ball Games," J. Royal Stat. Soc. A (1971), p. 623.

[14] Selby, S.M., Standard Mathematical Tables (14th edition), Chemical Rubber Company (1965).

[15] Soeroso, ., "The Asymptotic Efficiency of the Moment Estimators for Gamma Distribution Parameters," Masters Thesis, USNPS, March 1976.

[16] Sokolnikoff, I.S., Advanced Calculus, McGraw-Hill, 1939.

INITIAL DISTRIBUTION LIST

	No. Copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
Dean of Research Code 023 Naval Postgraduate School Monterey, California 93940	1
Library (Code 0212) Naval Postgraduate School Monterey, California 93940	2
Library (Code 55) Naval Postgraduate School Monterey, California 93940	2
Professor L. M. LeCam Department of Statistics University of California Berkeley, Ca. 94720	1
Department of Operations Research Naval Postgraduate School Monterey, Ca. 93940	
Attn: Professor D. R. Barr	1
Professor G. H. Bradley	1
Professor G. G. Brown	1
Professor J. Esary	1
Professor D. Gaver	1
Professor H.J. Larson	1
Professor P. Lewis	1
Professor K. Marshall	1
Professor P. Milch	1
Professor R. Read	20
Professor M. Sovereign	1
Professor P. Zehna	1
Professor Dan Davis Mathematics Department Naval Postgraduate School Monterey, California 93940	1

U177733

DUDLEY KNOX LIBRARY - RESEARCH REPORTS

5 6853 01067321 3